



Article

Integrating Thermal Infrared Imaging and Weather Data for Short-Term Prediction of Building Envelope Thermal Appearance

Nikolay Golosov ^{1,2,*} and Guido Cervone ^{1,2,3}

¹ Department of Geography, The Pennsylvania State University, University Park, PA 16802, USA; guc18@psu.edu

² Institute for Computational and Data Sciences (ICDS), The Pennsylvania State University, University Park, PA 16802, USA

³ Earth and Environmental Systems Institute (EESI), The Pennsylvania State University, University Park, PA 16802, USA

* Correspondence: golosov@psu.edu

Abstract: This study presents a novel deep-learning framework for predicting the thermal appearance of building envelopes under varying weather conditions based on a new dataset collected using a thermal infrared camera at 10 min intervals over a one-and-a-half-year period. Unlike existing studies that rely on simulated data or physical models that do not always accurately reflect the complex heat transfer processes in real buildings, we have collected a large dataset showing how a building behaves under different climatic conditions. We propose a novel deep-learning approach that integrates weather data and thermal imagery to predict the temperature distribution on the building façade for the next 24 and 48 h. The model uses a state-of-the-art recurrent neural network architecture, PredRNN V2, with an action conditioning mechanism to incorporate weather forecasting data into the prediction process. We evaluate this approach in terms of average accuracy, prediction accuracy in specific regions, and visual-perceptual performance of the images. The proposed framework achieves a prediction accuracy of 1.5 °C (root mean square error—RMSE) for the 24 h prediction and 2.04 °C (RMSE) for the 48 h prediction, outperforming baseline models in terms of temperature prediction accuracy and structural similarity of the predicted images.



Citation: Golosov, N.; Cervone, G. Integrating Thermal Infrared Imaging and Weather Data for Short-Term Prediction of Building Envelope Thermal Appearance. *Remote Sens.* **2024**, *16*, 3981. <https://doi.org/10.3390/rs16213981>

Academic Editor: Massimiliano Pepe

Received: 20 August 2024

Revised: 1 October 2024

Accepted: 23 October 2024

Published: 26 October 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: thermal remote sensing; deep learning; building envelope; temperature prediction; PredRNN; thermal infrared imaging; weather forecasting; urban heat transfer; energy efficiency

1. Introduction

Thermal imaging has emerged as a powerful tool for non-destructive evaluation of buildings, enabling the detection of heat losses, air leakages, and moisture issues, which can significantly affect energy efficiency and occupant comfort [1–3]. By capturing the infrared radiation emitted by building surfaces, thermal imaging can provide valuable insights into the thermal performance of building envelopes, helping to identify areas of poor insulation, thermal bridges, and other anomalies that may not be visible to the naked eye [4–6]. This information can be used to prioritize retrofit measures, optimize building designs, and monitor the effectiveness of energy conservation strategies [7,8].

The interpretation of thermal images is often hampered by the influence of environmental factors, such as ambient temperature, wind speed, and solar radiation [9,10]. These factors can cause significant variations in the surface temperatures of building elements, making it difficult to distinguish between genuine thermal anomalies and transient effects induced by the environment [1,2,4,9]. Emissivity is another important factor affecting the accuracy of temperature measurements using thermal imaging. It quantifies a material's ability to emit infrared energy relative to that of a perfect blackbody. Building materials

and surfaces exhibit significant variations in emissivity values [11], with distinct emissivity properties observed in glass windows and facade materials. Failure to account for these variations can lead to inaccuracies in temperature measurements [12]. To ensure accurate surface temperature evaluation in thermal imaging, one needs to consider the emissivity, reflected radiation, atmosphere self-emission and transmission which is affected by environmental factors such as air temperature, humidity, and the distance between the camera and the object. The change of these parameters throughout the day and across seasons necessitates their inclusion in radiance-to-temperature conversion calculation for more reliable surface temperature assessments [13].

Several studies have investigated the use of numerical models to predict the thermal behavior of building envelopes under varying environmental conditions [14–16]. These models typically integrate thermal imaging data with meteorological measurements and building material properties to simulate the heat transfer processes occurring within the building envelope [4,15]. The accuracy of these physics-based models is often limited by the computation of the heat transfer phenomena and the uncertainties associated with the input parameters [1,4,8].

The accurate prediction of thermal behavior in building surfaces has implications beyond individual building performance, extending to urban climate studies. Urban areas, which are characterized by dense concentrations of buildings and infrastructure, experience unique thermal dynamics that contribute to phenomena like the urban heat island effect [17]. The utilization of small-scale weather prediction models enables better understanding and mitigation of the impacts of urbanization on local climates [18]. These predictive capabilities can inform urban planning decisions, enhance energy management strategies, and contribute to the development of more sustainable and resilient cities [19]. By leveraging advanced modeling techniques and considering the complex interactions between buildings and their environment, researchers can provide valuable insights into the thermal behavior of urban landscapes, ultimately guiding efforts to create more livable and climate-resilient urban spaces.

Recent advancements in deep learning have significantly enhanced the capabilities of building energy simulation by providing more accurate predictions of temperature distribution in buildings [20–23]. Convolutional neural networks (CNNs) have been particularly successful in imaging applications due to their ability to capture spatial patterns and hierarchical features in image data [24]. CNNs have been used for various thermal imaging tasks, such as defect detection [25–27], energy efficiency assessment [28,29], and thermal comfort prediction [30]. On the other hand, the application of deep learning to thermal image prediction in the context of building performance assessment is still in its early stages. Existing studies often use synthetic data [31], focus on predicting one-dimensional temperature time series rather than 2D images [32,33], or predict temperatures at large spatiotemporal scales that do not capture building-level details [34,35]. The integration of multi-modal data such as weather forecasts and thermal images into deep learning models remains a challenge, as these data are often heterogeneous and vary in terms of temporal and spatial resolutions [36,37]. Meteorological variables, such as ambient temperature, wind speed, and solar radiation, have a significant impact on the thermal behavior of buildings, but they are typically measured at different locations and time intervals than the thermal images [8]. To effectively utilize these heterogeneous data sources, deep learning models need to be designed with appropriate data fusion and preprocessing techniques [38].

To address these limitations, we propose a novel approach for thermal image prediction that combines a state-of-the-art recurrent neural network (RNN) architecture with weather forecasting data. Specifically, we adopt the PredRNN V2 model [39], which has shown superior performance in video prediction tasks, and extend it to accommodate the unique characteristics of thermal image sequences and meteorological data. To incorporate the effects of environmental factors on the thermal behavior of the building envelope, we consider the environmental factors as drivers of change in the spatial appearance of the thermal images. To account for these external factors, modifying the image appearance,

we use the action conditioning mechanism that fuses the weather forecasting data with the thermal image features at each time step [39]. This approach enables the model to adapt its predictions based on the expected changes in the ambient conditions, such as temperature fluctuations, wind speed variations, and solar radiation intensity. The main novelty of our approach lies in the combination of a state-of-the-art RNN architecture with an action conditioning mechanism that integrates weather forecasting data. It is designed to generate accurate thermal image predictions over extended time horizons (24 to 48 h) by leveraging the temporal modeling capabilities of the PredRNN V2 model and the contextual information provided by the weather forecast.

To evaluate the proposed framework, we collected a large-scale dataset of thermal images and weather measurements, covering a wide range of environmental conditions, collected over 456 days (about 1 year and 3 months). We compare the performance of the proposed framework against a baseline model that relies solely on thermal image data, demonstrating the benefits of incorporating weather forecasting information into the prediction process.

The main contributions of this study can be summarized as follows:

- We present a deep learning framework for thermal image prediction that integrates a state-of-the-art RNN architecture (PredRNN V2) with weather forecasting data, enabling accurate prediction of temperature distribution in the 2D thermal images over the two time horizons (24 to 48 h).
- We introduce the use of an action conditioning mechanism that fuses the thermal image features with the meteorological data at each time step, allowing the model to adapt its predictions based on the expected changes in environmental conditions.
- We evaluate the proposed framework using a large-scale dataset of thermal images collected over 456 days (about 1 year and 3 months). The dataset covers a wide range of environmental conditions, ensuring the generalizability of the results.
- We compare the performance of the proposed approach against two baseline models: naïve prediction that considers the previous day to be the next day and a baseline model that relies solely on thermal image data, demonstrating the benefits of incorporating weather forecasting information into the prediction process.

This paper is organized as follows: Section 1 provides an overview of the related work on thermal image prediction and deep learning applications in building performance assessment. Section 2 describes the dataset used in this study. Section 3 discusses the proposed methodology, including the PredRNN V2 architecture, the action conditioning mechanism and accuracy assessment. Section 4 presents the experimental results and discusses the performance of the proposed framework under different prediction scenarios. Finally, Section 5 concludes the paper and outlines future research directions.

2. Data

2.1. Thermal Images

The thermal infrared images for this study were collected using the Flir A70 thermal camera, manufactured by FLIR Systems, Inc., Wilsonville, OR, USA. The camera was installed on the roof of the Walker building. The camera consists of two sensors—thermal infrared and a visible light sensor. The thermal infrared sensor operates in a 7.5–14.0 μm wavelength range with a spatial resolution of 640×480 pixels and a temperature range of -20 to 175 $^{\circ}\text{C}$ and provides a field of view of 29° . The accuracy of the sensor is $\pm 2\%$ or 2 $^{\circ}\text{C}$. A visible light sensor provides a spatial resolution of 1920×960 pixels. The camera is located 148 m away from the target, which results in a pixel size of 11.9 cm. The camera was installed in a protective housing designed to shield it from direct sunlight, rain, and snow. The thermal imaging setup is shown in Figure 1.

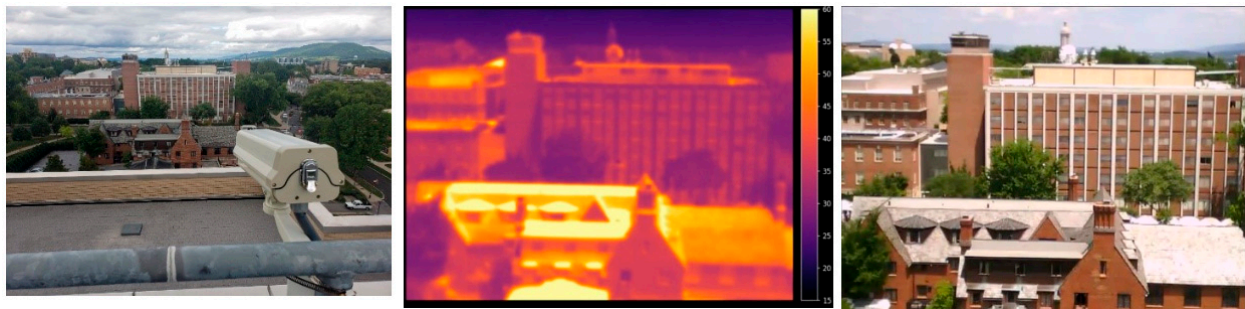


Figure 1. Thermal imaging setup and example infrared and visible light images captured by the FLIR A70 thermal imaging camera.

The camera was installed on the roof of the Walker building at location $40^{\circ}47'35.9''\text{N}$ $77^{\circ}52'00.4''\text{W}$, and oriented NE with azimuth 43.06° . It generally looks towards the Deike building and collects thermal and visible light images every 10 min. The camera is equipped with an HTTP server that was used to retrieve thermal and visible-light images for further processing. Thermal image data were collected using the Flir Research Studio, version 2.1.0 software, while the visible-light imagery was collected using the open-source FFMPEG, version 2022-07-06 software. The visible-light imagery was not used in this study, but we made it available on the data-sharing website (see the data availability statement). The captured thermal images had a resolution of 640×480 pixels, with each pixel representing a temperature value in degrees Celsius. Due to a technical limitation of the data collection software, we used a single emissivity value of 0.95 to calculate the temperature values. The thermal image dataset used in this study was collected over 456 days (about 1 year and 3 months), spanning from 15 July 2022 to 6 November 2023. A total of 65,664 thermal images were captured during this period, covering a wide range of environmental conditions. Images corresponding to different types of weather conditions are represented in Figure 2.

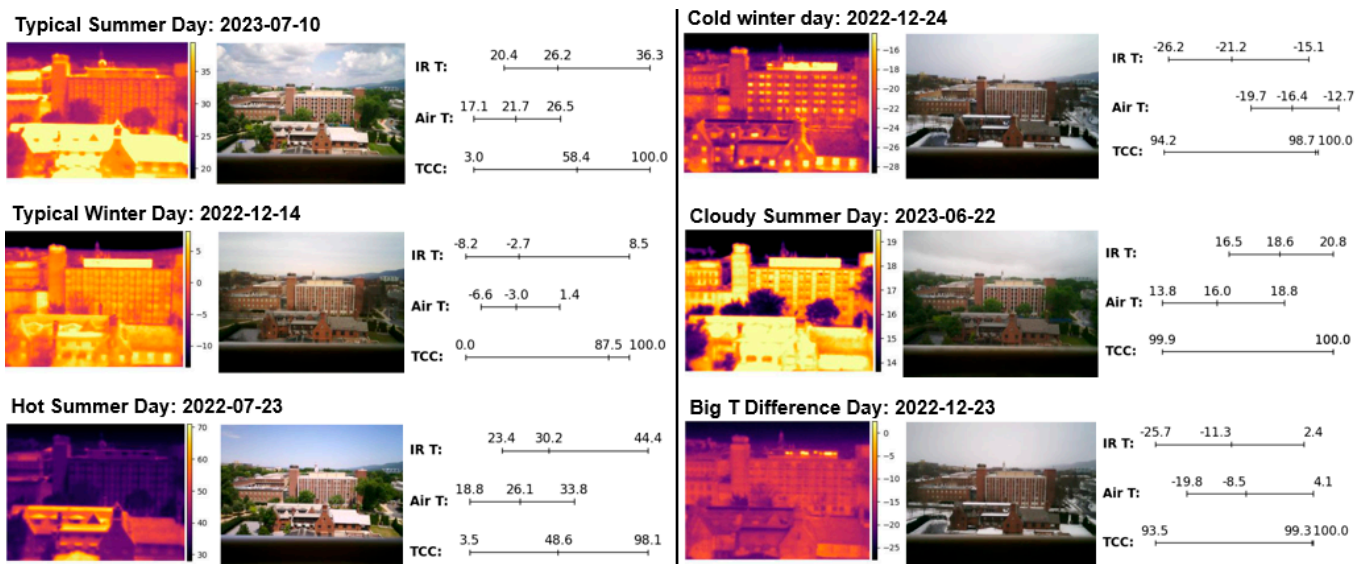


Figure 2. Representative thermal and visible light images for different weather conditions. Infrared images (color represents temperature distribution), visible-light images and their corresponding meteorological data are shown. Number bars represent minimum–average–maximum values of the meteorological variable. The variables are abbreviated as follows: IR T—average infrared image temperature (°C); Air T—air temperature at 2 m (°C); and TCC—total cloud cover (%).

To create a representative subset of thermal images for model training and evaluation, the preprocessed images were grouped into hourly intervals, and the average temperature values were computed for each pixel. This resulted in a total of 10,944 hourly thermal images, with each image representing the average temperature distribution over one hour. To reduce the computational complexity and memory requirements of the deep learning model, the representative subset of the building façade with a resolution of 64×64 pixels, containing the windows, walls, and ventilation latices and unobstructed with vegetation, was subset from the entire image. The example image and the subset configuration are presented in Figure 3.

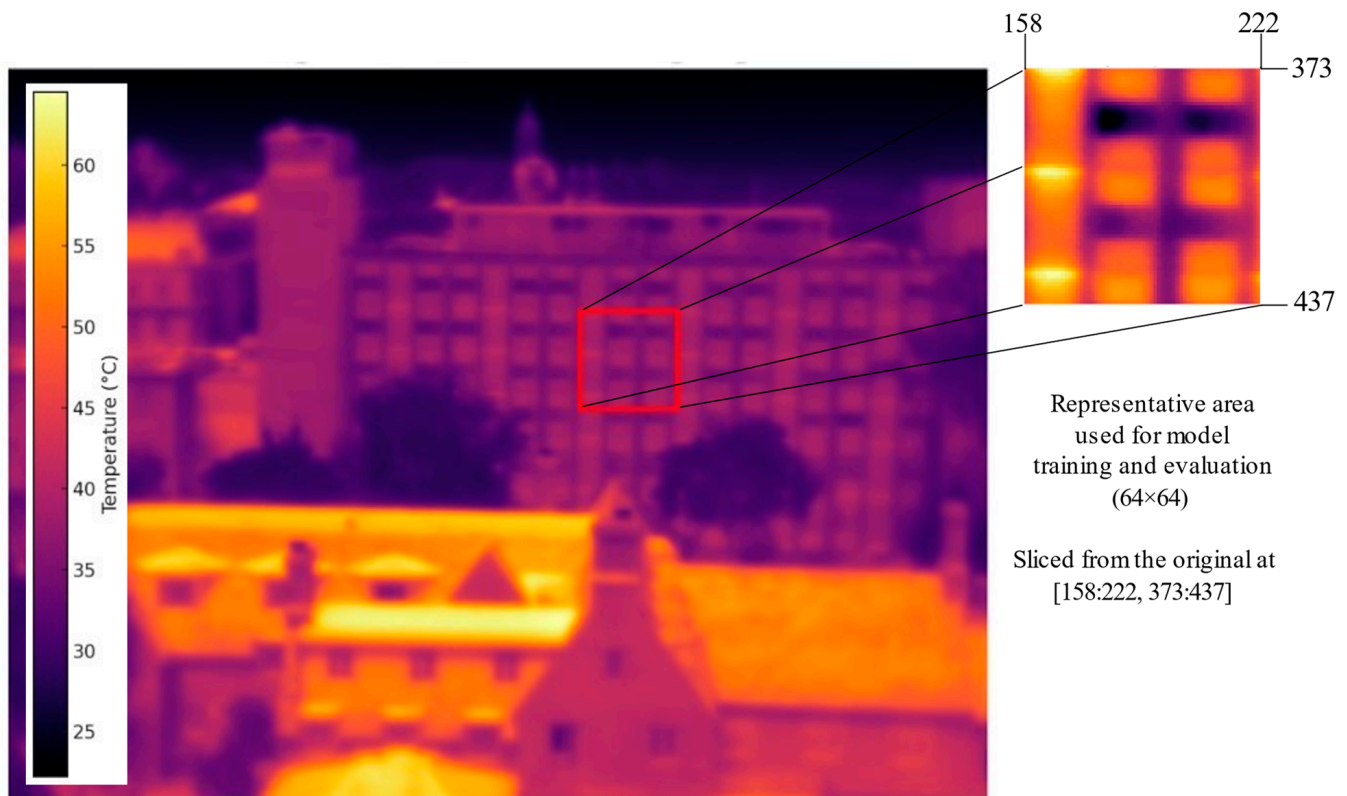


Figure 3. A representative subset of the thermal image is used for training and evaluation.

The preprocessed images were then normalized to a range of $[0, 1]$ by dividing each pixel value by the maximum temperature in the dataset ($55.03\text{ }^{\circ}\text{C}$). The preprocessed images were then split into training and test sets, with a ratio of 80% and 20%, respectively. The training set was used to optimize the parameters of the deep learning model, while the unseen testing set was used to assess the model's performance. The test set was used to evaluate the final performance of the trained model and compare it with the baseline approach.

2.2. Weather Data

To incorporate the effects of environmental factors on the thermal behavior of the building envelope, weather data were collected from the ERA5 reanalysis dataset [37]. ERA5 is a global atmospheric reanalysis dataset produced by the European Centre for Medium-Range Weather Forecasts (ECMWF), providing hourly estimates of various meteorological variables at a spatial resolution of $0.25^{\circ} \times 0.25^{\circ}$ [37].

The following meteorological variables were extracted from the ERA5 dataset for the location of the building site:

- The 2 m temperature (t2m): The air temperature at 2 m above the surface, in degrees Celsius.
- The 10 m U-component of wind (u10): The eastward component of the wind speed at a height of 10 m above the surface, in meters per second.
- The 10 m V-component of wind (v10): The northward component of the wind speed at a height of 10 m above the surface, in meters per second.
- The 2 m dewpoint temperature (d2m): The temperature at which the air must be cooled to reach saturation, in degrees Celsius.
- Surface solar radiation (ssr): The amount of solar radiation reaching the Earth's surface, in watts per square meter.
- Total cloud cover (tcc): The fraction of the sky covered by clouds, as a percentage.
- Total precipitation (tp): The accumulated liquid and frozen water, including rain and snow, in meters.
- Sun elevation: the elevation of the sun above the horizon in degrees, calculated at the geographic location of the Deike building.

These variables were selected based on their relevance to the thermal behavior of building envelopes and their availability in the ERA5 dataset. The wind components (u10 and v10) were used to inform the model of the wind speed and direction, which influence the convective heat transfer on the building's surface [9]. The dewpoint temperature (d2m) was used to estimate the relative humidity, which affects the latent heat transfer and moisture content of the building materials [5].

The meteorological variables were extracted for the same time period as the thermal image dataset (15 July 2022, to 6 November 2023) using the Copernicus Data Store toolbox. The extracted weather variables were then joined by time with the hourly thermal images, resulting in a complete dataset of 10,944 samples, each containing a thermal image and the corresponding weather variables. To ensure the compatibility of the weather data with the deep learning model, the meteorological variables were normalized to a range of [0, 1] using a min–max scaling. The minimum and maximum values of each variable were computed from the training set and used to scale the values in the validation and test sets.

The normalized weather data were then concatenated with the thermal image data to form a multi-modal input for the deep learning model. The data input into the model were formatted as training sequences, each containing a $48 \times 64 \times 64 \times 9$ tensor, and saved as a separate NumPy .npz file. The first dimension of the tensor, 48, represents the number of images and accompanying weather parameters, consisting of 24 images and parameters from the previous day, followed by 24 images and parameters from the next day. The second and third dimensions, 64×64 , represent the dimensions of the input single-channel thermal images. The last dimension consists of the image data in the first channel and the accompanying weather variables in the last eight channels. All the data are scaled to the range [0, 1]. An example of the input training sequences is provided in Figure 4.

The dataset used in this study consists of 10,944 hourly thermal images and corresponding weather variables collected over 456 days (about 1 year and 3 months). The thermal images were preprocessed and subset to a resolution of 64×64 pixels, while the weather data were extracted from the ERA5 reanalysis dataset at the tile coordinates covering the building site location. The thermal and weather data were normalized and concatenated to form a multi-modal input for the deep learning model. To evaluate the prediction models, the dataset was split into training and test sets, with a ratio of 80% and 20%, respectively.

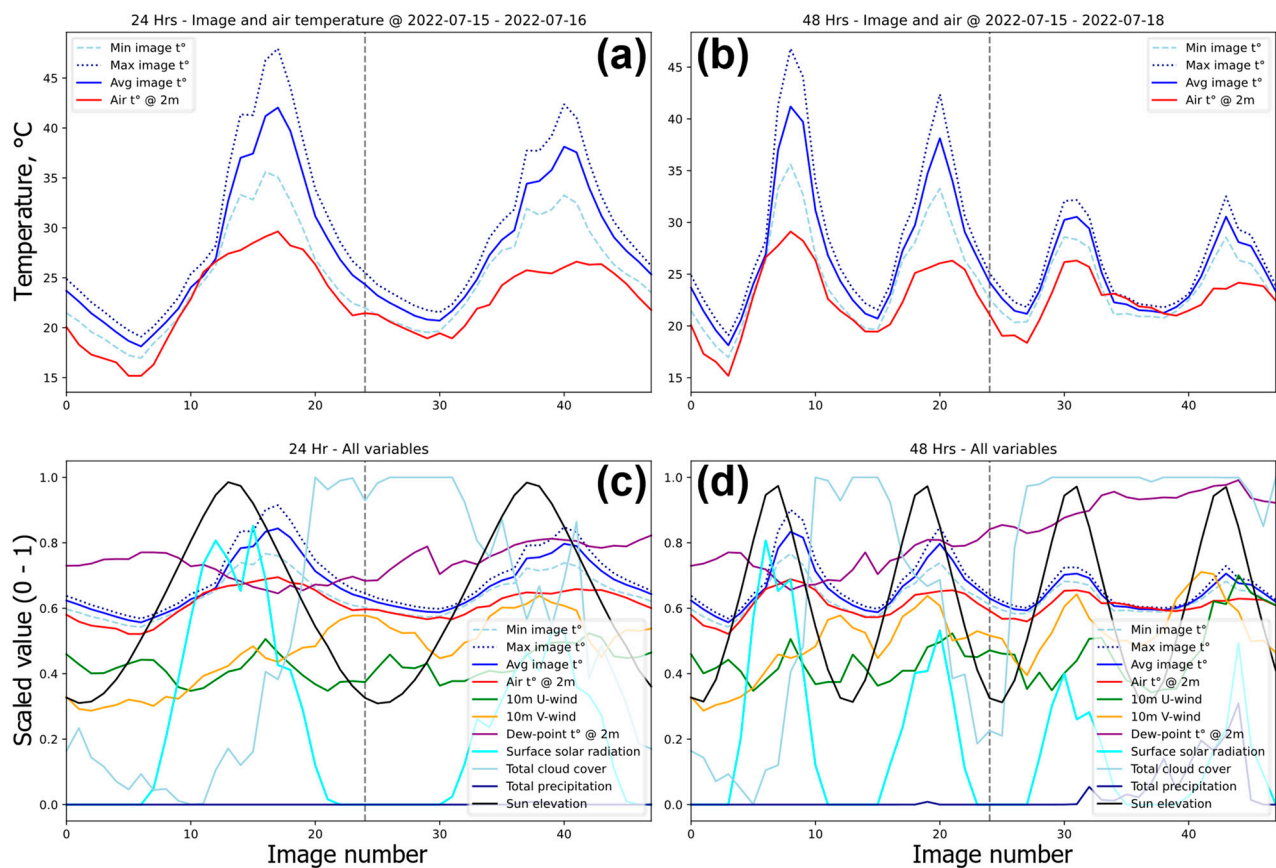


Figure 4. Examples of the training data. Air and image temperature inputs: (a) 24 h of the previous day followed by 24 h of next-day data. The average–min–max image temperature of the 64×64 array is shown accompanied by EPA5 air temperature; (b) 48 h followed by 48 h. Weather variable and image temperature inputs: (c) 24 h of the previous day /24 next-day data accompanied by the EPA5 weather variables. The data are min–max scaled between [0, 1]; (d) 48 h of the previous period followed by 48 h of images and weather data from the next period. We have used one image per hour in the 24 h prediction models and one image per two hours in the 48 h prediction models.

3. Methods

3.1. General Considerations

The study consisted of four main phases: data preprocessing, sequence generation, model training, and model evaluation. Figure 5 outlines these steps in more detail.

In the preprocessing phase, thermal images were processed to generate hourly average images, while ERA5 NetCDF data were used to extract weather variables at specific longitude and latitude coordinates. Concurrently, the sun elevation at each hour in the dataset was calculated and the weather data were consolidated in a look-up table. The sequence generation phase involved creating different datasets for training and testing purposes to gain insights into the contribution of weather data to the accuracy of thermal image prediction. The following training scenarios were evaluated (see Figure 6):

1. Images only, no weather (NoW): A baseline model that does not contain any weather variables.
2. Images with observed temperature (ActW): A model where the images are accompanied by the actual air temperature.
3. Forecasted temperature (Fcast): Similar to ActW but using forecasted temperature instead of actual image temperature.

4. Forecasted temperature, same image (Fcast-Si): A model to gain insight into the contribution of image data to prediction performance. A single repeated image is used as input, accompanied by forecasted temperature.

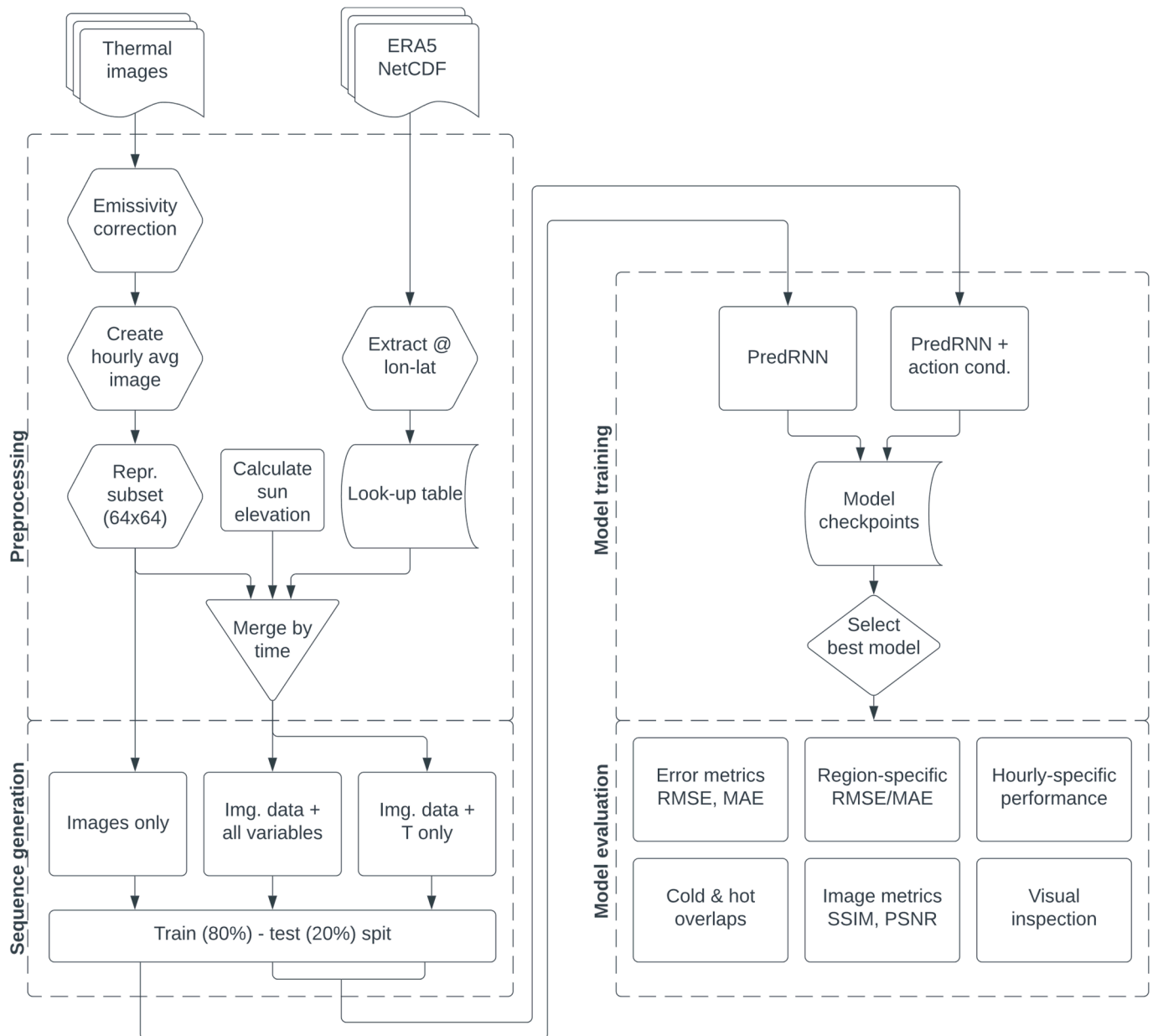


Figure 5. Study workflow diagram illustrating the preprocessing, training data generation, model training, and evaluation stages.

For model training, two variations of the PredRNN algorithm were employed: a standard PredRNN for modeling the scenario without weather variables and a PredRNN with an action conditioning mechanism to model the contribution of weather variables as drivers dictating the spatiotemporal changes in the images. These models were iteratively trained with checkpointing to capture intermediate states. Post-training, the best-performing model was selected based on the lowest RMSE (root mean squared error) between the ground truth and predicted temperatures, evaluated on the unseen test data.

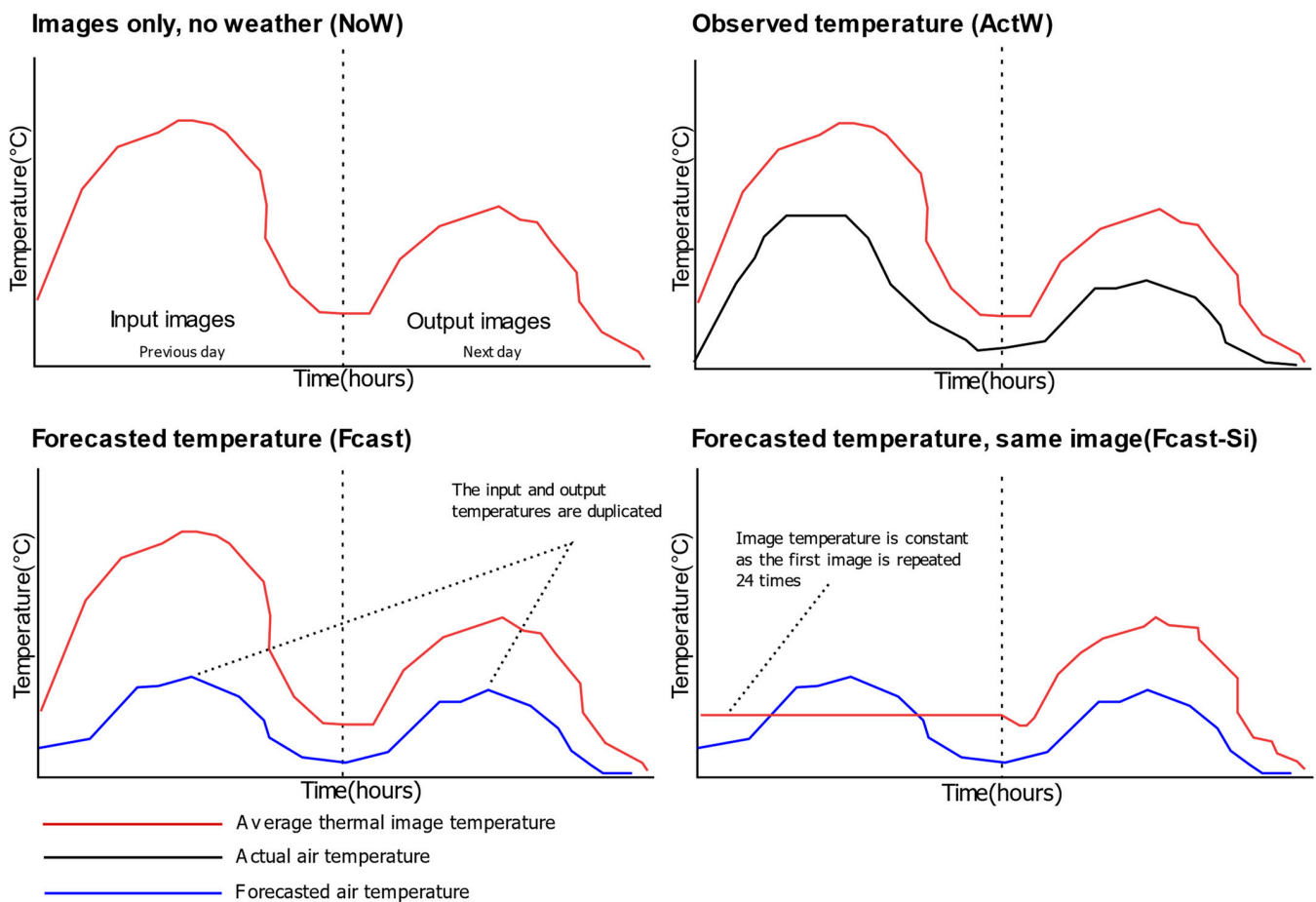


Figure 6. Training scenarios.

The model evaluation phase encompassed several metrics and methods to ensure comprehensive assessment. These included RMSE and MAE (mean average error) for error measurement, both global and region-specific; cold and hot spot overlap analysis; image quality metrics such as SSIM (structural similarity index metric) and PSNR (peak signal-to-noise ratio); and visual inspection for qualitative assessment.

The proposed thermal image prediction framework is based on the PredRNN V2 architecture [39], a state-of-the-art recurrent neural network (RNN) designed for spatiotemporal predictive learning. The PredRNN V2 model extends the original PredRNN architecture [40] by introducing a hierarchical structure of convolutional long short-term memory (ConvLSTM) units [41] and a gradient highway unit (GHU) for efficient gradient propagation [40].

The core building block of the PredRNN V2 model is the ConvLSTM unit, which combines the spatial feature extraction capabilities of convolutional neural networks (CNNs) with the temporal modeling capabilities of long short-term memory (LSTM) units [36]. The ConvLSTM unit replaces the fully connected layers in the traditional LSTM with convolutional layers, enabling the model to capture both the spatial and temporal dependencies in the input data [41]. The PredRNN V2 model consists of multiple stacked ConvLSTM units, organized in a hierarchical structure [40]. The hierarchical structure allows the model to learn representations at different spatial and temporal scales, capturing both the short-term and long-term dependencies in the thermal image sequence. The lower layers of the model focus on extracting local spatial features, while the higher layers capture the global temporal dynamics [40].

The input to the PredRNN V2 model is a sequence of thermal images, denoted as $X = X_1, X_2, \dots, X_T$, where X_t is the thermal image at time step t , and T is the length of the

input sequence. The model processes the input sequence through a series of ConvLSTM units, generating a hidden state representation at each time step [40]. The hidden state at time step t , denoted as H_t , is computed as follows:

$$\begin{aligned} I_t &= \sigma(W_{xi} * X_t + W_{hi} * H_{t-1} + b_i) \\ F_t &= \sigma(W_{xf} * X_t + W_{hf} * H_{t-1} + b_f) \\ C_t &= F_t \odot C_{t-1} + I_t \odot \tanh(W_{xc} * X_t + W_{hc} * H_{t-1} + b_c) \\ O_t &= \sigma(W_{xo} * X_t + W_{ho} * H_{t-1} + b_o) \\ H_t &= O_t \odot \tanh(C_t) \end{aligned} \quad (1)$$

where I_t , F_t , C_t , and O_t are the input gate, forget gate, cell state, and output gate, respectively; W and W_h are the convolutional weights for the input and hidden state; b represents the bias terms; σ is the sigmoid activation function; $\odot$ denotes the element-wise multiplication; and $*$ denotes the convolution operation [41].

To facilitate the flow of gradients through the deep hierarchical structure, the PredRNN V2 model incorporates a gradient highway unit (GHU) between the ConvLSTM layers [40]. The GHU is designed to adaptively combine the hidden states from the current and previous layers, allowing the gradients to bypass the intermediate layers if needed. This mechanism helps to alleviate the vanishing gradient problem and enables the model to learn long-term dependencies more effectively [40].

The output of the PredRNN V2 model is a sequence of predicted thermal images, denoted as $Y = Y_1, Y_2, \dots, Y_{T'}$, where Y_t is the predicted thermal image at time step t , and T' is the length of the output sequence. The predicted thermal images are generated by applying a convolutional layer to the hidden state at each time step, followed by a sigmoid activation function to map the output values to the range $[0, 1]$ [40].

3.2. Action Conditioning Mechanism

To incorporate the effects of environmental factors on the thermal behavior of the building envelope, we introduce an action conditioning mechanism that fuses the weather forecasting data with the thermal image features at each time step [40]. The action conditioning mechanism allows the model to adapt its predictions based on the expected changes in the ambient conditions, such as temperature fluctuations, wind speed variations, and solar radiation intensity.

The weather forecasting data are represented as a sequence of weather vectors, denoted as $W = W_1, W_2, \dots, W_T$, where W_t is the weather vector at time step t , containing the normalized values of the meteorological variables (e.g., temperature, wind speed, and solar radiation). The weather vectors are concatenated with the thermal image features along the channel dimension, forming a multi-modal input for the PredRNN V2 model [40].

The action conditioning mechanism modifies the computation of the hidden state in the ConvLSTM units by incorporating the weather vector at each time step [40]. Specifically, the input gate, forget gate and cell state equations are updated as follows:

$$\begin{aligned} I_t &= \sigma(W_{xi} * X_t + W_{hi} * H_{t-1} + W_{wi} * W_t + b_i) \\ F_t &= \sigma(W_{xf} * X_t + W_{hf} * H_{t-1} + W_{wf} * W_t + b_f) \\ C_t &= F_t \odot C_{t-1} + I_t \odot \tanh(W_{xc} * X_t + W_{hc} * H_{t-1} + W_{wc} * W_t + b_c) \end{aligned} \quad (2)$$

where W_{wi} , W_{wf} , and W_{wc} are the convolutional weights for the weather vector for the input gate, forget gate, and cell state equations respectively, and the other terms remain the same as in the original ConvLSTM equations [40].

By conditioning the hidden state computation on the weather vector, the PredRNN V2 model can learn to adjust its predictions based on the expected environmental conditions. For example, if the weather forecast indicates a high solar radiation intensity, the model can learn to predict higher surface temperatures on the building facade, considering the additional heat gain from the sun [8].

Figure 7 demonstrates the network architecture diagram. The input data of the $1 \times 48 \times 64 \times 64 \times 9$ contain 48 input images of shape 64×64 . There are 24 input and 24 output images (48 total). During the training, the input consists of 24 + 24 images. During the inference, the output images are not used, and they are masked. There are nine channels in the input data. The first channel is a thermal image, and the remaining eight channels are actioning conditioning data, i.e., weather variables. The 1d weather value is repeated to fill the 64×64 layer by the same value. The output of the model is a shape of $1 \times 48 \times 64 \times 64 \times 1$, where the first 24 frames are input images and 24 images are predicted images. The software code for this study was implemented in Python programming language 3.10.13 using the PyTorch library version 1.12.2. Training and evaluation of the proposed framework was performed on the computer equipped with a single Nvidia RTX A4500 graphics card which has 20 GB of video memory (VRAM) and is manufactured by Nvidia Corporation, located in Santa Clara, CA, USA.

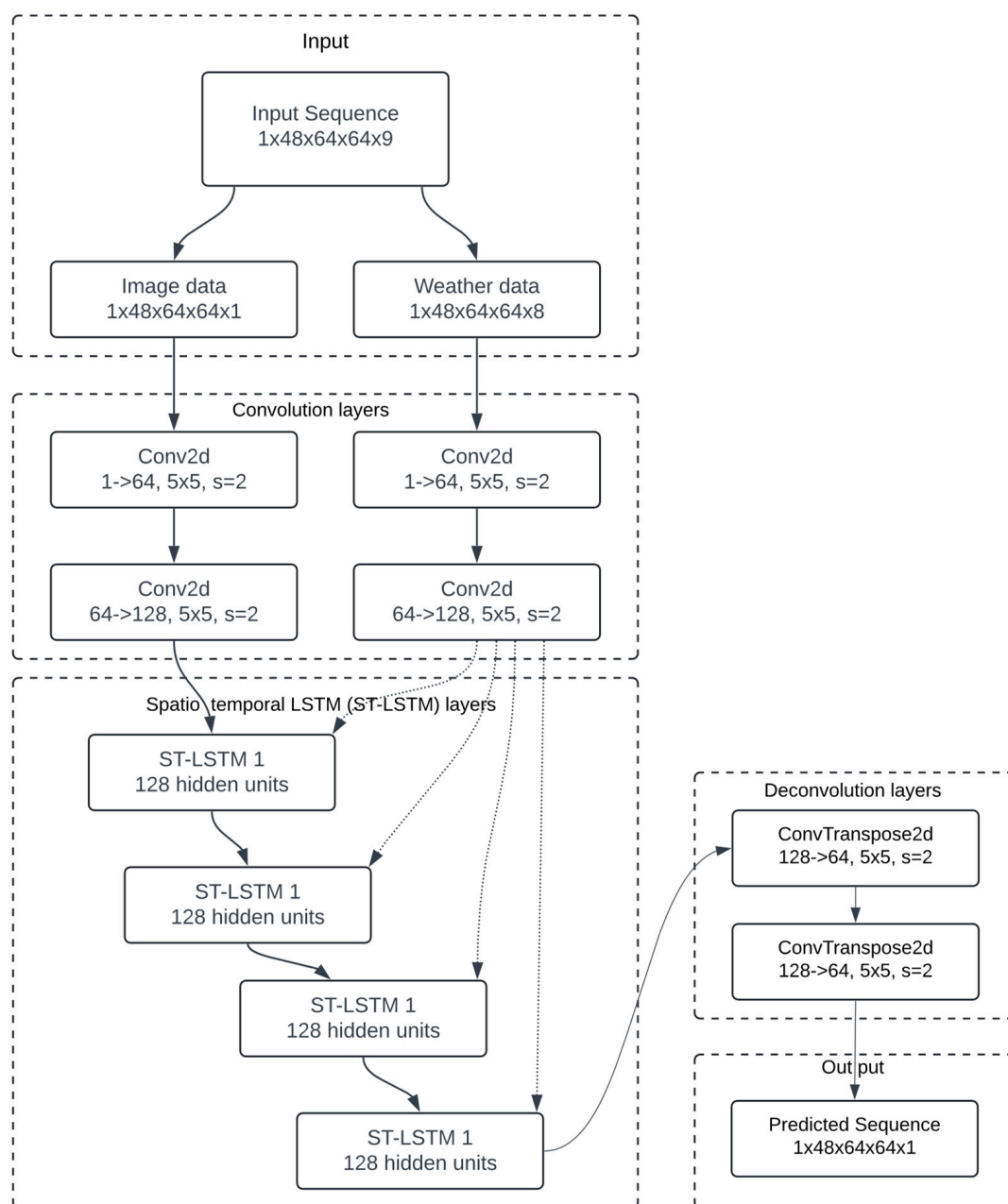


Figure 7. Network diagram of the PredRNN network with action conditioning.

3.3. Training and Evaluation

The PredRNN V2 model with action conditioning is trained using a supervised learning approach, with the objective of minimizing the mean squared error (MSE) between the predicted and ground-truth thermal images) [40]. The MSE loss function is computed as follows:

$$L_{MSE} = \frac{1}{T' \cdot N} \sum_{t=1}^{T'} \sum_{i=1}^N (Y_{t_i} - \hat{Y}_{t_i})^2 \quad (3)$$

where Y_{t_i} is the ground-truth temperature value at pixel i and time step t , $\hat{Y}_{t_i}$ is the predicted temperature value, T' is the length of the output sequence, and N is the total number of pixels in the thermal image [40].

The model is trained using the Adam optimizer [42] with a learning rate of 0.0005, which, in experiments, was found to avoid the loss explosion that occurred with the default learning rate of 0.001 and a batch size of 16. The models were trained to 160,000 iterations. The models were checkpointed every 5000 iterations, and the most accurate model was selected by evaluating each checkpoint's performance on the test data.

To assess the performance of the models, we employed a diverse set of evaluation metrics. The model is evaluated on the test set, which consists of thermal image sequences and corresponding weather data that are not used during training. These metrics were calculated on a per-hour basis and averaged over all the test files for each model. Further, to assess the effectiveness of the action conditioning mechanism, we compare the performance of the proposed framework with a baseline model that relies solely on thermal image data without incorporating the weather forecasting information. The baseline model has the same PredRNN V2 architecture as the proposed framework but does not include the action conditioning mechanism.

To evaluate the accuracy of the temperature prediction in the image, we use two metrics: the mean absolute error (MAE) and the root mean squared error (RMSE) [43]. The MAE measures the average absolute difference between the predicted and ground-truth temperature values, while the RMSE measures the average squared difference, giving more weight to large errors [44].

3.4. Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE)

We calculated the RMSE and MAE between the predicted and ground truth frames to measure the pixel-level accuracy of the predictions. These metrics provide an overall assessment of the model's ability to accurately predict temperature values. The RMSE is calculated as follows:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (4)$$

where y_i is the ground truth value, $\hat{y}_i$ is the predicted value, and n is the total number of pixels.

The MAE is calculated as follows:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (5)$$

3.5. Cold and Hot Spot Overlap and IoU

We introduced novel metrics to evaluate the model's performance in predicting extreme temperature regions. Cold and hot spots were defined based on temperature percentiles, and we calculated the overlap and intersection over union (IoU) between the predicted and ground truth spots [45]. The overlap is calculated as follows:

$$Overlap = \frac{|P \cap G|}{|G|} \quad (6)$$

where P is the set of predicted spots and G is the set of ground truth spots. The IoU is calculated as follows:

$$IoU = \frac{|P \cap G|}{|P \cup G|} \quad (7)$$

These metrics gauge the model's ability to capture the spatial patterns of extreme temperatures accurately.

3.6. Polygon-Based MAE and RMSE

We defined specific polygonal regions of interest within the frames, representing cold (windows) and hot areas (brick wall areas). These regions were manually outlined using the daylight visual image of the building, which was captured by a second (RGB) sensor of the Flir A70 camera. The location of the polygonal regions used in this evaluation step is illustrated in Figure 8.

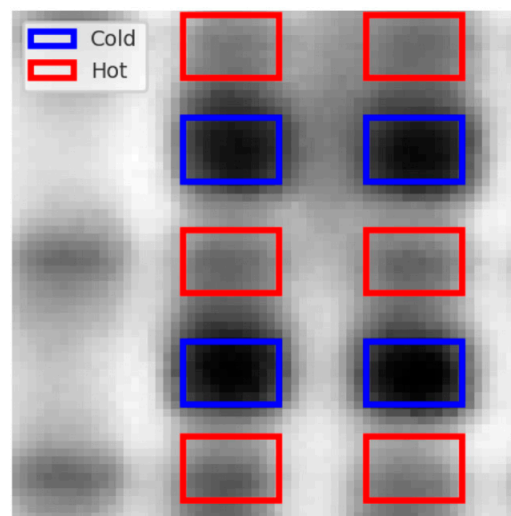


Figure 8. Location of the cold (windows) and hot (walls) polygonal regions for calculating polygon-based MAE and RMSE.

The MAE and RMSE were calculated for the predicted and ground truth values within these regions using the same formulas as described in point 1. This analysis provides insights into the model's accuracy in predicting temperatures in critical areas of interest.

3.7. Structural Similarity Index Measure (SSIM) and Peak Signal-to-Noise Ratio (PSNR)

To assess the perceptual quality of the predictions, we utilized SSIM [46] and PSNR [47]. SSIM measures the structural similarity between two images and is calculated as follows:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (8)$$

where μ_x and μ_y are the means, σ_x^2 and σ_y^2 are the variances, and σ_{xy} is the covariance of images x and y . c_1 and c_2 are kept as small constants to avoid instability.

PSNR measures the ratio between the maximum possible power of a signal and the power of corrupting noise. It is calculated as follows:

$$PSNR = 10\log_{10}\left(\frac{MAX_I^2}{MSE}\right) \quad (9)$$

where MAX_I is the maximum possible pixel value of the image and MSE is the mean squared error between the two images.

Higher values of SSIM and PSNR indicate better perceptual quality of the predictions.

The evaluation process was repeated for each model checkpoint and test file, accumulating the per-hour metrics and overall metrics to compare across different model architectures and configurations.

In addition to the quantitative evaluation, we also perform a qualitative analysis of the predicted thermal images, comparing them with the ground-truth images to assess the visual quality and consistency of the predictions. Qualitative analysis helps to identify any systematic errors or artifacts in the predictions and provides insights into the model's ability to capture the spatial and temporal patterns in the thermal image sequences.

4. Results

4.1. Prediction Performance

The proposed deep learning framework, based on the PredRNN V2 model with action conditioning, was evaluated on the test set of thermal image sequences and corresponding weather data. The model's performance was assessed using multiple metrics: mean absolute error (MAE) and root mean squared error (RMSE). We used region-specific metrics as well as SSIM and PSNR.

For the 24 h prediction horizon, the model incorporating all weather variables achieved the lowest MAE and RMSE, outperforming the naive “previous-day” prediction in 46 out of 46 test cases. The model without weather variables exhibited the worst accuracy, while the models with temperature data alone showed intermediate performance (Figure 9).

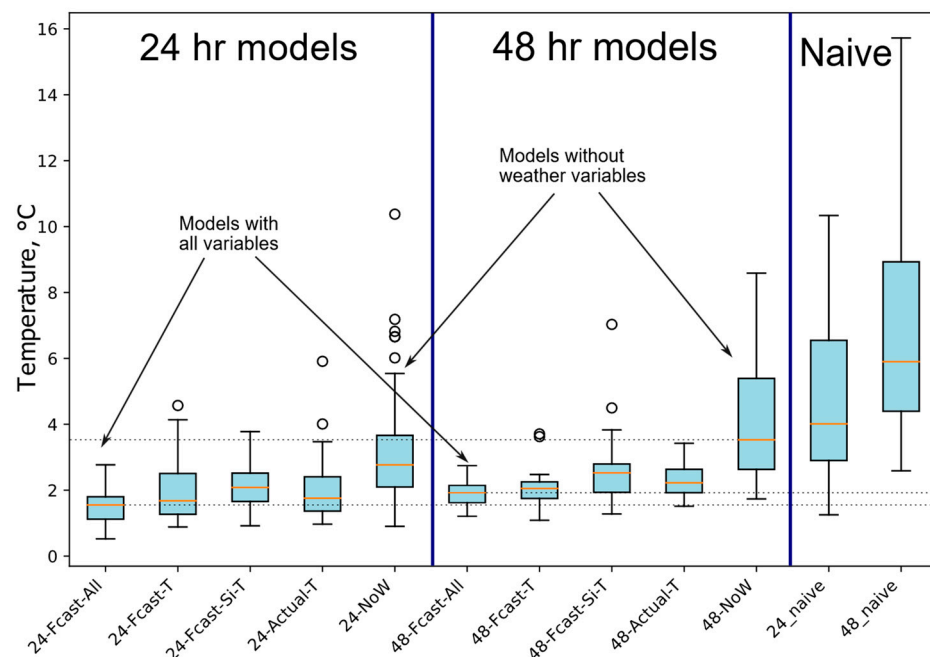


Figure 9. Distribution of errors across the test cases for different prediction models.

Similarly, for the 48 h prediction horizon, the model with all weather variables demonstrated the best performance, followed by the models with temperature data and the model without weather variables. The overall accuracy across the 24 h prediction horizon was higher compared to the 48 h horizon, as expected due to the increased uncertainty over longer periods. In terms of average RMSE, all the models outperformed the naive prediction (labeled as 24_naive and 48_naive in the plots) and models without weather variables (labeled as 24-NoW and 48-NoW) (Figure 9).

4.2. Error Distribution

The distribution of MAE and RMSE values across the test cases was analyzed to assess the consistency of the model's performance. For both the 24 h and 48 h prediction horizons, the model incorporating all weather variables showed the lowest spread of errors, indicating more reliable predictions (Figure 9). In contrast, the model without weather variables exhibited a wider range of errors, highlighting the importance of incorporating environmental factors into the prediction process.

4.3. Temporal Error Distribution

To investigate the temporal variation of prediction errors, the average MAE and RMSE values were computed for each hour of the prediction horizon across all test cases. The results showed that the model's performance varied depending on the time of day, with larger errors occurring during the daytime hours (Figure 10). This pattern was consistent across all models, suggesting that the increased solar radiation and temperature fluctuations during the day pose challenges for accurate thermal image prediction.

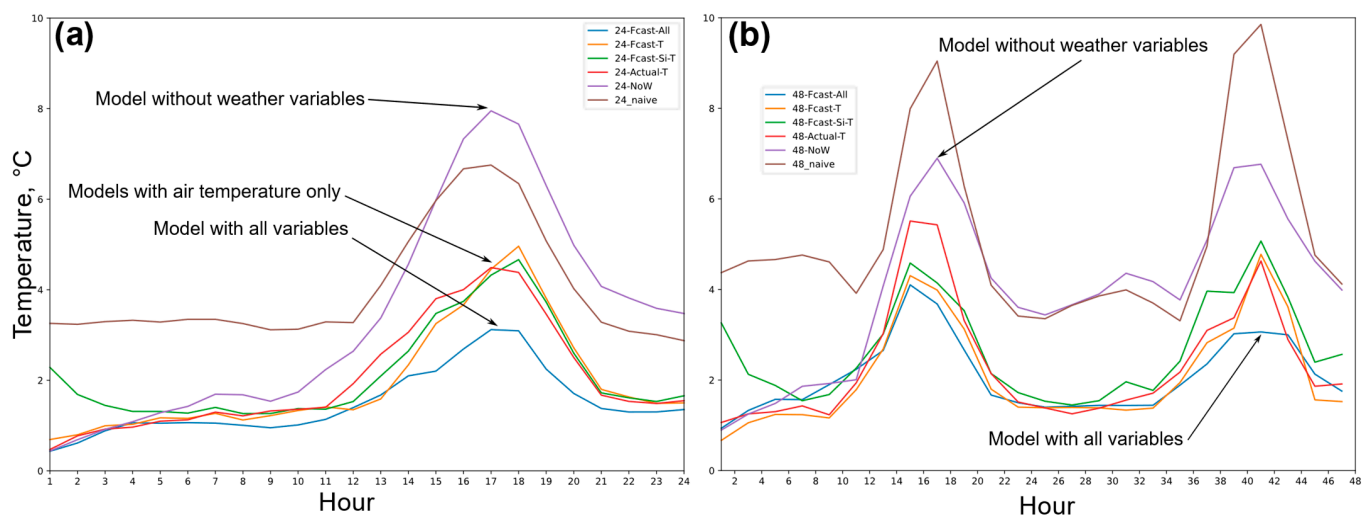


Figure 10. Average error (RMSE) per hour for the 24 h models (a) and for 48 h models (b).

The model incorporating all weather variables consistently achieved the lowest hourly MAE and RMSE values, demonstrating its ability to adapt to changing environmental conditions. This is true for both 24 and 48 h prediction horizons. The model with forecasted temperature data showed slightly lower errors (RMSE of 1.95 °C) compared to the model using actual weather data (RMSE of 2.02 °C) (Figure 10).

The comparison of the spread of the error by hour across the test cases for the 24 and 48 h models is presented in Figure 11.

Comparing the 24 h models, there is a considerable reduction in the spread of errors across all hours when weather variables are included (b) compared to the model without weather data (a). This indicates that incorporating weather information significantly improves the consistency and accuracy of the 24 h temperature predictions. Similarly, for the 48 h models, including weather variables (d) noticeably reduces the spread of errors compared to the model without weather data (c). However, the overall spread of errors is generally higher for the 48 h model with weather variables compared to the 24 h model with weather variables. This suggests that while weather data help to improve the 48 h model's accuracy, the inherent uncertainty in longer-range forecasts still leads to greater variability in errors compared to the 24 h model. Interestingly, when comparing the models without weather variables (a and c), there is no significant visible difference in the spread of errors between the 24 h and 48 h temperature predictions. This implies that the inclusion

of weather data is the primary factor driving the improvement in accuracy and reduction in error spread, particularly for the 24 h model.

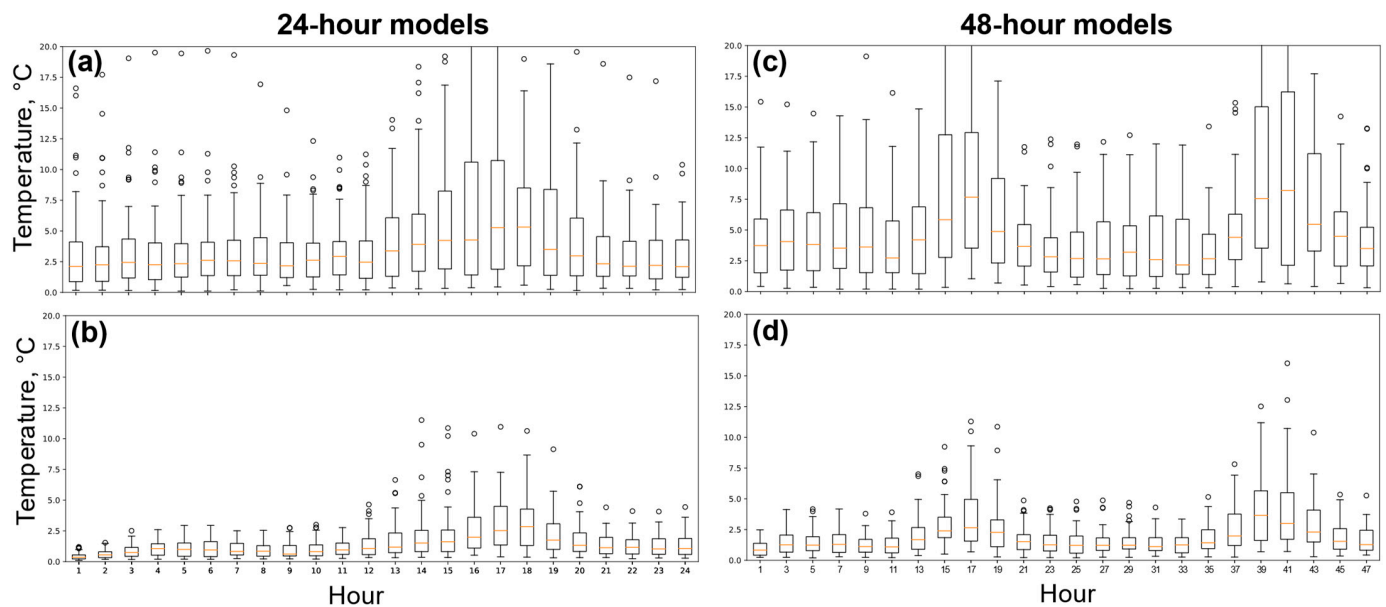


Figure 11. Spread of the error by hour for the 24 and 48 h models: (a) 24 h model without weather variables (24-NoW); (b) 24 h model with all weather variables included (24-Fcast-All); (c) 48 h model without weather variables (48-NoW) (d) 48 h model with all weather variables included (48-Fcast-All).

4.4. Comparison of 24 h and 48 h Predictions

The root mean square error (RMSE) values exhibited a consistent increase when comparing the 24 h and 48 h prediction horizons across all models (Table 1).

Table 1. Comparison of metrics across the selected models.

Model Name	RMSE	MAE	Cold Ovr	Cold IoU	Hot Ovr	Hot IoU	Cold RMSE	Hot RMSE	SSIM	PSNR
24 h prediction										
Naïve model	4.77	3.91	0.60	0.48	0.60	0.48	3.91	4.42	0.60	15.10
24-Fcast-All *	1.49	1.41	0.62	0.49	0.61	0.47	1.37	1.44	0.69	11.70
24-Fcast-T	1.95	1.81	0.44	0.31	0.41	0.27	1.75	1.89	0.39	11.80
24-Fcast-Si-T	2.14	1.97	0.39	0.26	0.34	0.22	1.88	2.04	0.31	10.71
24-Actual-T	2.02	1.87	0.41	0.28	0.41	0.28	1.79	1.94	0.38	11.61
24-NoW	3.35	3.27	0.57	0.44	0.52	0.39	3.00	3.31	0.59	6.42
48 h prediction										
Naïve model	6.55	5.11	0.51	0.39	0.51	0.40	4.79	5.43	0.55	12.47
48-Fcast-All *	2.02	1.95	0.57	0.44	0.57	0.43	1.88	1.95	0.67	8.63
48-Fcast-T	2.11	1.98	0.45	0.32	0.39	0.26	1.86	2.02	0.48	10.08
48-Fcast-Si-T	2.68	2.40	0.29	0.18	0.26	0.16	2.42	2.58	0.21	12.69
48-Actual-T	2.35	2.14	0.31	0.20	0.27	0.17	2.11	2.26	0.23	12.39
48-NoW	4.01	3.89	0.42	0.30	0.36	0.25	3.60	3.97	0.37	5.12

* denotes the best performing model.

The naive model showed the most substantial increase in RMSE, from 4.77 for the 24 h prediction to 6.55 for the 48 h prediction. Similarly, the Fcast-All model, which achieved the lowest RMSE among the 24 h models at 1.49, saw an increase to 2.02 for the 48 h prediction. The Fcast-T, Fcast-Si-T, and Actual-T models also experienced higher RMSE values for the 48 h prediction, with the Fcast-Si-T model showing the most significant increase from 2.14 to 2.68. The model without weather (NoW)'s RMSE rose from 3.35 for

the 24 h prediction to 4.01 for the 48 h prediction. These findings underscore the growing challenge of maintaining prediction accuracy as the forecast horizon extends, likely due to the accumulation of errors and the increasing influence of unobserved factors over longer periods.

4.5. Region-Specific Performance

The prediction performance was further analyzed for specific regions of interest within the thermal images, such as windows (cold RMSE) and walls (hot RMSE). The cold RMSE values were consistently lower than the hot RMSE values across all models and prediction horizons (Table 1). For the 24 h prediction, the cold RMSE ranged from 1.37 (24-Fcast-All) to 3.91 (naive), while the hot RMSE ranged from 1.44 (24-Fcast-All) to 4.42 (Naive). Similarly, for the 48 h prediction, the cold RMSE varied from 1.88 (48-Fcast-All) to 4.79 (Naive), and the hot RMSE ranged from 1.95 (48-Fcast-All) to 5.43 (naive). The lower cold RMSE values for windows can be attributed to the more stable thermal behavior of glazing materials compared to walls. Windows have lower thermal mass and are less influenced by the outdoor environment's thermal fluctuations. In contrast, the higher hot RMSE values for walls can be explained by the complex heat transfer processes occurring within the wall structure. Walls have higher thermal mass and are more sensitive to changes in outdoor temperature, solar radiation, and wind conditions. The thermal inertia of walls leads to a delayed response to external factors, making accurate predictions more challenging, particularly over longer time horizons. The model incorporating all weather variables demonstrated the best performance for both windows and walls, indicating its ability to capture the distinct thermal characteristics of different building components. In particular, the inclusion of solar radiation data contributed to the improved prediction accuracy for windows, as it helped to account for the direct solar heat gain through the glazed surfaces.

Structural Similarity Index (SSIM) comparison: The Structural Similarity Index (SSIM) was employed to assess the models' ability to capture the structural information and perceptual quality of the predicted thermal images. Higher SSIM values indicate better structural similarity between the predicted and ground truth images. The 24 h prediction models generally exhibited higher SSIM values compared to their 48 h counterparts (Table 1). Among the 24 h models, the Fcast-All and NoW models achieved the highest SSIM values of 0.69 and 0.59, respectively, while the Fcast-Si-T model had the lowest SSIM of 0.31. For the 48 h prediction, the Fcast-All model maintained a high SSIM value of 0.67, whereas the Fcast-Si-T and Actual-T models showed a significant drop in SSIM to 0.21 and 0.23, respectively. The higher SSIM values of the Fcast-All models suggest that incorporating a comprehensive set of weather variables or relying solely on the thermal image sequence can help preserve the structural information in the predicted images. The lower SSIM values of the Fcast-Si-T and Actual-T models, particularly for the 48 h prediction, indicate that using a single repeated thermal image or actual temperature data may not be sufficient to capture the complex spatial patterns and textures of the thermal images over longer time horizons.

PSNR comparison: The PSNR values varied significantly among the models and prediction horizons. Higher PSNR values indicate better image quality and less noise in the predictions. The Fcast-All model, despite its superior performance in RMSE, MAE, and SSIM, had relatively low PSNR values. This can be attributed to its ability to capture more detailed and complex patterns in the thermal images by incorporating a comprehensive set of weather variables, which may introduce some noise or small variations in the predicted pixel values. In contrast, the Fcast-Si-T and Actual-T models, which rely on simpler inputs, produced smoother and less noisy predictions, resulting in higher PSNR values. These findings suggest a trade-off between capturing fine details and maintaining a high signal-to-noise ratio in thermal image prediction, highlighting the importance of considering multiple evaluation metrics and application-specific requirements when selecting the most appropriate model.

An overview of the predicted thermal images generated by the two best-performing models, one using all weather variables forecasted for 24 h (24-Fcast-All) and the other using all variables forecasted for 48 h (48-Fcast-All), is provided in Figure 12.

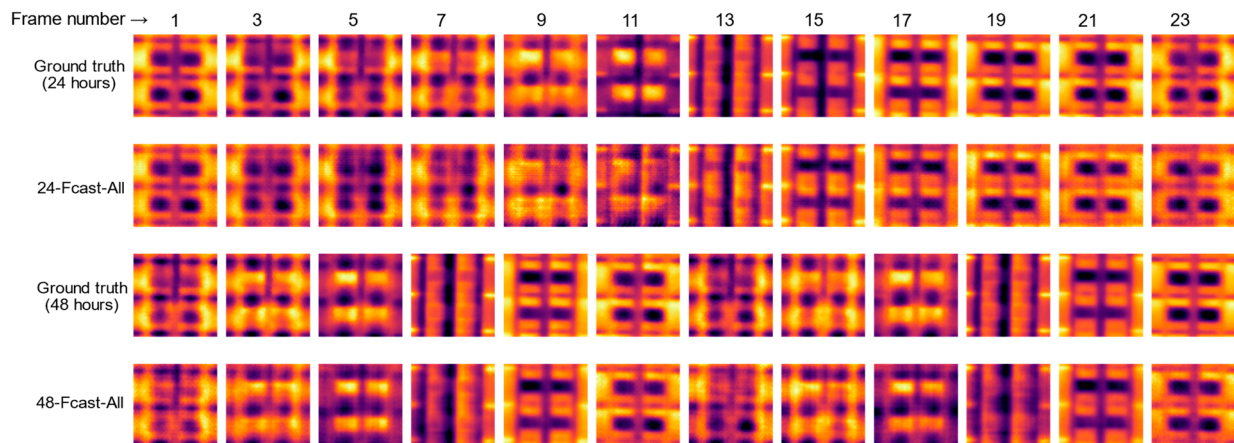


Figure 12. Examples of prediction by the selected 24 and 48 h models.

The predicted thermal images generally resemble their corresponding ground truth images, capturing the overall spatial patterns and temperature distributions. However, some fine details and small-scale features are not predicted with perfect accuracy. For instance, the predicted image for Frame 11 in the 24-Fcast-All model exhibits noticeable differences compared to its ground truth counterpart. Additionally, the predicted thermal images have a slightly more grainy and pixelated appearance than the ground truth images, which is reflected in the relatively low Peak Signal-to-Noise Ratio (PSNR) values reported in Table 1. Despite these limitations, the models demonstrate a promising ability to forecast the general patterns and trends of temperature distributions over the 24 and 48 h time horizons.

4.6. Comparison of the Model Performance Under Different Scenarios

To further understand the model prediction behavior under different scenarios, we selected four different inputs from the test set, as depicted in Figure 13. The first scenario, labeled as (a) in the diagram, shows two mild winter days on 11–12 December 2022. There is a small change (about 2.5 °C) between the days in the input IR image temperature. The change in forecasted ambient air temperature is also small and accurately predicted in the weather forecast. The model without access to forecasted weather variables (labeled as “Img. only” in the diagram) significantly overpredicted the maximum temperature by about 15 °C. This model, as evident from the graphs, expects the temperature to rise and fall during the day. It can make small adjustments in its prediction, but they mainly affect the magnitude of the maximum value. However, if the day is without a significant rise and fall in temperature, the model cannot produce a “flat line” prediction—there will always be a spike, varying in its magnitude. Visual inspection of the images shows that the temperature in the output images of the model without weather variables (labeled as “Img. only”) rapidly rises, as expressed by the bright yellow color in the image in column $t = 17$. The spatial features in the nighttime and early morning images ($t = 1, 5, 9$) appear relatively close to the ground truth images and their temperature distribution. However, in the daytime images ($t = 13$ to 21), the temperature appears too high, and most spatial details are lost.

The next scenario, labeled as (b), is like the previous one and shows a relatively cold spring day on 11–12 March 2023, with about a 6 °C difference in temperature. The model with weather variables adjusted its prediction to be close to the air temperature forecast. There is some more evident loss of spatial details in the daylight images ($t = 13$ to 21), but the results are still much closer to the ground truth compared to the model without weather variables. Notably, the model without weather variables misinterprets the small spike in temperature from the previous day as a signal to further increase the predicted temperature compared to scenario (a).

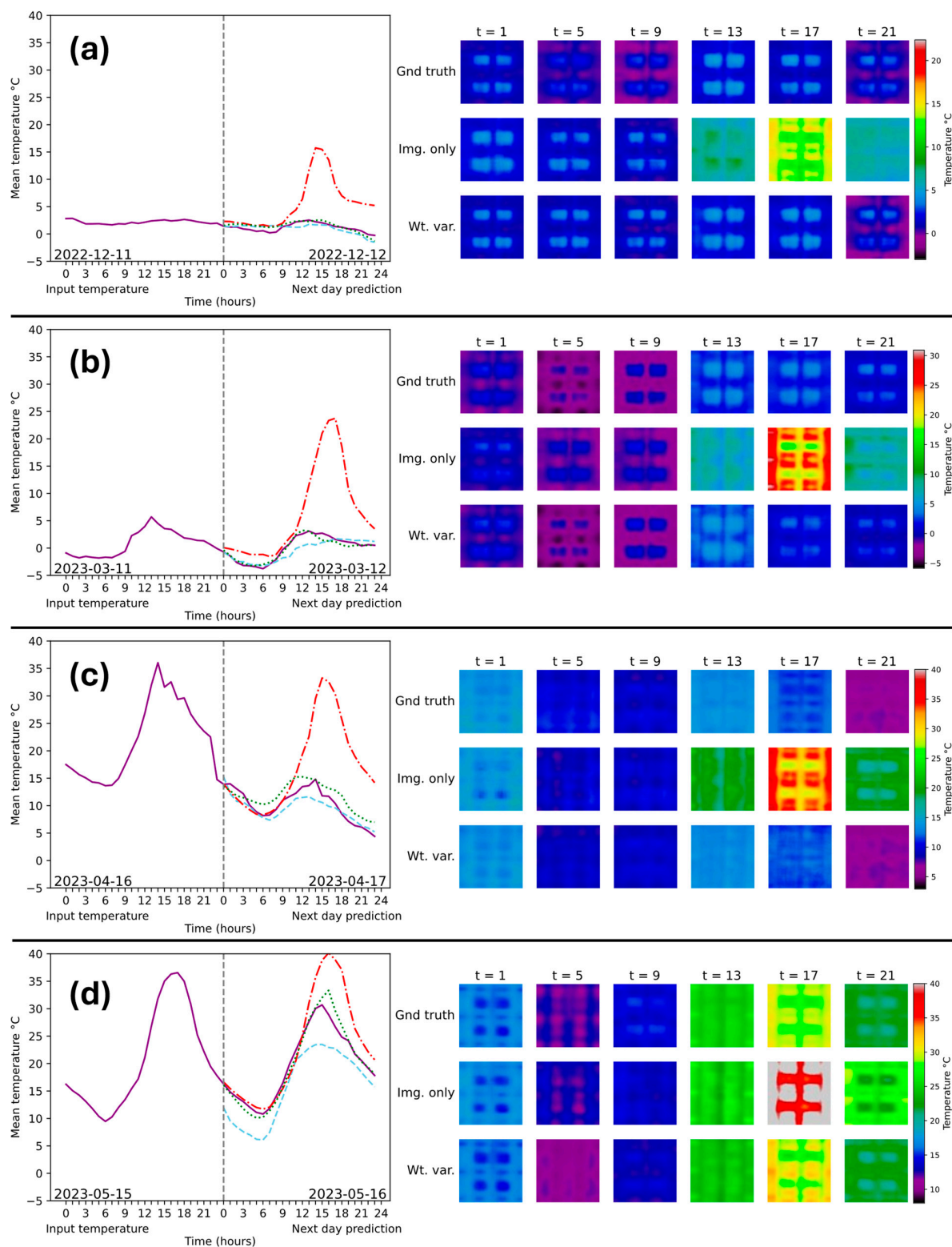


Figure 13. Comparison of the performance of thermal image prediction models with and without weather forecast data for four different scenarios: (a) two mild winter days without a significant change in temperature, (b) relatively cold spring days with a small change in temperature, (c) a hot spring day followed by a significantly cooler day, and (d) two warm spring days. Two models are compared against the ground-truth thermal images. The abbreviations used are “Gnd truth/IR Image T” for ground truth and input IR image temperature, “Img. only/Image only pred.” for the model using only thermal images, and “Wt. var/Weather var. pred.” for the model incorporating weather forecast variables.

Scenario (c) depicts a hot spring day on 16 March and a cooler day on 17 March 2023, exhibiting a rapid rise in temperature on the first day, while the next day's temperature fell significantly, which was accurately captured by the weather forecast. The model with access to the weather forecast successfully adjusted its predictions and produced a reasonable spatial appearance of the image. The model without weather variables produced sharply rising and declining average temperature curves, as in the previous scenarios.

Finally, scenario (d) shows warm spring days on 15 and 16 May 2023. The change in temperature generally followed bell curves on both days, with the second day being colder than the first. The model without weather variables overpredicted the temperature, while the model with access to the weather forecast produced a relatively accurate prediction. The graph shows that the air temperature in the forecast does not rise as sharply as the thermal image temperature, but it appears that the signal indicating the thermal image temperature will rise more rapidly than the air temperature was captured in the other weather variables.

In conclusion, the model with access to all weather variables produces a more reasonable forecast of the thermal appearance of the façade compared to the model without access to weather variables. The model without access to weather variables tends to significantly overpredict the temperature when the next day is too different from the previous day.

5. Discussion

The results of this study demonstrate the effectiveness of the proposed deep learning framework for thermal image prediction, which combines the PredRNN V2 model with an action conditioning mechanism to incorporate weather forecasting data. The model incorporating all weather variables consistently outperformed the baseline models, highlighting the importance of considering environmental factors when predicting the thermal behavior of building envelopes.

The superior performance of the model with all weather variables can be attributed to its ability to capture the complex interactions between the building facade and the ambient conditions. By conditioning the predictions on the expected weather patterns, the model can adapt to the varying solar radiation intensity, wind speed, and precipitation levels, which significantly influence the surface temperatures and heat transfer processes.

The temporal error analysis revealed that the prediction accuracy varies depending on the time of day, with larger errors occurring during the daytime hours. This finding suggests that additional research is needed to improve the model's performance under challenging daytime conditions, potentially by incorporating more detailed solar radiation data or adaptive modeling techniques that can capture the rapid temperature fluctuations caused by changing solar angles and cloud cover.

The comparison of the 24 h and 48 h prediction models highlighted the increasing difficulty of maintaining accurate predictions over longer time horizons. This observation underscores the need for further research into long-term thermal image prediction, focusing on strategies to mitigate the accumulation of errors and to capture the slower dynamics of heat transfer within the building envelope.

The region-specific performance analysis demonstrated the model's ability to capture the distinct thermal characteristics of different building components, such as windows and walls. The inclusion of solar radiation data proved particularly beneficial for predicting the thermal behavior of windows, which are more sensitive to shortwave radiation compared to opaque wall surfaces. This finding highlights the potential of the proposed framework for component-level thermal performance assessment and targeted retrofitting strategies.

The perceptual quality assessment using SSIM and PSNR metrics highlighted the trade-off between capturing fine details and maintaining a high signal-to-noise ratio in thermal image prediction. While the model incorporating all weather variables achieved superior performance in terms of prediction accuracy, it also introduced some noise or small variations in the predicted pixel values. This finding suggests that the choice of the

most appropriate model should consider the specific application requirements and the desired balance between prediction accuracy and image quality.

Limitations of the Study

Our study has several limitations that should be considered when interpreting the results. Firstly, due to a technical limitation of the data collection software, we used a single emissivity value of 0.95. The representative subset of the building consisted of two materials—red bricks and glass—with different emissivity values (often-used values are 0.93 and 0.92–0.94), leading to an error in temperature estimates. One solution could be to use a wall subset without windows and a single emissivity value. However, this approach, despite yielding more accurate temperature values in the physical sense, would not allow us to evaluate the spatial predictive skills of the model due to the lack of spatial features on a homogeneous subset of the building wall. Such a task could be performed by one-dimensional time series prediction algorithms (such as Facebook Prophet [48]), which is beyond the scope of this article. In future studies, with an improved measurement setup, we plan to address this inconsistency by measuring the actual kinetic temperature of the building materials to calculate the actual emissivity values.

Another limitation is that atmospheric parameters, such as humidity and air temperature, vary across different days and seasons, affecting the transmissivity of the atmosphere between the sensor and the camera. These changes could be accounted for with a more complex measurement setup that measures radiance in energy units, such as the commonly used $W/m^2/\mu m/sr$, to enable the atmosphere and emissivity contribution to be accounted for more accurately, instead of recording the camera-calculated temperature. Furthermore, since building materials are not perfect emitters, with emissivity values below 1, it would be beneficial to account for the reflected infrared radiation and atmospheric self-emission, which contribute to the measured signal and affect the recorded temperature values.

Finally, the dataset is limited to a single building, restricting the set of observed building materials and climatic conditions. To enhance the generalizability and robustness of the proposed framework, future research should focus on expanding the dataset to include a wider range of building types, climatic conditions, and seasonal variations. Additionally, the integration of high-resolution weather forecasting models and real-time sensor data could further improve the prediction accuracy and enable near-real-time thermal performance monitoring.

6. Conclusions

This study presents a novel deep learning framework for thermal image prediction that integrates a state-of-the-art PredRNN V2 model with an action conditioning mechanism to incorporate weather forecasting data. The proposed approach aims to enhance the accuracy and reliability of thermal image prediction over extended time horizons, enabling better-informed decision-making in building performance assessment and energy efficiency optimization.

The experimental results show the proposed framework's effectiveness, with the model incorporating all weather variables consistently outperforming the baseline models across various prediction horizons and performance metrics. The inclusion of environmental factors, such as solar radiation, wind speed, and precipitation levels, proved crucial for capturing the complex interactions between the building facade and the ambient conditions, leading to more accurate and adaptable predictions.

The temporal error analysis revealed variations in prediction accuracy depending on the time of day, with larger errors occurring during the daytime hours. This finding highlights the challenges associated with predicting the rapid temperature fluctuations caused by changing solar angles and cloud cover, emphasizing the need for further research into adaptive modeling techniques and the incorporation of more detailed solar radiation data.

The comparison of the 24 h and 48 h prediction models underscored the increasing difficulty of maintaining accurate predictions over longer time horizons, as the accumu-

lation of errors and the influence of unobserved factors become more significant. This observation calls for future research to focus on strategies for mitigating error propagation and capturing the slower dynamics of heat transfer within the building envelope.

The region-specific performance analysis demonstrated the model's ability to capture the distinct thermal characteristics of different building components, such as windows and walls. This finding offers possibilities for component-level thermal performance assessment and targeted retrofitting strategies, enabling more efficient and cost-effective building energy optimization.

The proposed deep learning framework for thermal image prediction demonstrates promising results, but several limitations should be addressed in future research. The current study is based on a dataset covering a specific set of building types and climatic conditions and expanding the dataset to include a wider range of building geometries, construction materials, and climatic zones would enhance the generalizability and robustness of the approach. The use of ERA5 reanalysis data with a grid cell resolution of 0.25° (31 km) may not fully capture local microclimate variations, and integrating high-resolution weather forecasting models and on-site sensor data could improve the accuracy and localization of the predictions. The lack of interpretability in the model may hinder its adoption in practice, and exploring techniques for enhancing the interpretability and explainability of the model is crucial.

In conclusion, the proposed deep learning framework demonstrates its effectiveness for thermal image prediction. By integrating recurrent neural networks with weather forecasting data, the proposed approach improves the precision of the temperature distribution predictions.

Author Contributions: Conceptualization, N.G. and G.C.; methodology, N.G.; software, N.G.; validation, N.G. and G.C.; formal analysis, N.G.; investigation, N.G.; resources, G.C.; data curation, N.G.; writing—original draft preparation, N.G.; writing—review and editing, G.C.; visualization, N.G.; supervision, G.C.; project administration, G.C.; funding acquisition, G.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received partial funding from the Sustainability Fund of the College of Earth and Mineral Sciences at the Pennsylvania State University.

Data Availability Statement: The data, including thermal infrared images, used in this study are available at https://prosecco.geog.psu.edu/inferno_web/ (accessed on 19 August 2024).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Fox, M.; Coley, D.; Goodhew, S.; De Wilde, P. Thermography Methodologies for Detecting Energy Related Building Defects. *Renew. Sustain. Energy Rev.* **2014**, *40*, 296–310. [\[CrossRef\]](#)
2. Kirimtat, A.; Krejcar, O. A Review of Infrared Thermography for the Investigation of Building Envelopes: Advances and Prospects. *Energy Build.* **2018**, *176*, 390–406. [\[CrossRef\]](#)
3. Mayer, Z.; Epperlein, A.; Vollmer, E.; Volk, R.; Schultmann, F. Investigating the Quality of UAV-Based Images for the Thermographic Analysis of Buildings. *Remote Sens.* **2023**, *15*, 301. [\[CrossRef\]](#)
4. Kylili, A.; Fokaides, P.A.; Christou, P.; Kalogirou, S.A. Infrared Thermography (IRT) Applications for Building Diagnostics: A Review. *Appl. Energy* **2014**, *134*, 531–549. [\[CrossRef\]](#)
5. Lucchi, E. Applications of the Infrared Thermography in the Energy Audit of Buildings: A Review. *Renew. Sustain. Energy Rev.* **2018**, *82*, 3077–3090. [\[CrossRef\]](#)
6. Brooke, C. Thermal Imaging for the Archaeological Investigation of Historic Buildings. *Remote Sens.* **2018**, *10*, 1401. [\[CrossRef\]](#)
7. Barreira, E.; de Freitas, V.P. Evaluation of Building Materials Using Infrared Thermography. *Constr. Build. Mater.* **2007**, *21*, 218–224. [\[CrossRef\]](#)
8. Nardi, I.; Lucchi, E.; de Rubeis, T.; Ambrosini, D. Quantification of Heat Energy Losses through the Building Envelope: A State-of-the-Art Analysis with Critical and Comprehensive Review on Infrared Thermography. *Build. Environ.* **2018**, *146*, 190–205. [\[CrossRef\]](#)
9. Lehmann, B.; Wakili, K.G.; Frank, T.; Collado, B.V.; Tanner, C. Effects of Individual Climatic Parameters on the Infrared Thermography of Buildings. *Appl. Energy* **2013**, *110*, 29–43. [\[CrossRef\]](#)

10. Olbina, S.; Beliveau, Y. Developing a Transparent Shading Device as a Daylighting System. *Build. Res. Inf.* **2009**, *37*, 148–163. [CrossRef]
11. Maldague, X. *Theory and Practice of Infrared Technology for Nondestructive Testing*; Wiley series in microwave and optical engineering; Wiley: New York, NY, USA, 2001; ISBN 978-0-471-18190-3.
12. Madding, R.P. Emissivity Measurement and Temperature Correction Accuracy Considerations. In Proceedings of the SPIE, Thermosense XXI, Orlando, FL, USA, 5–9 April 1999; Volume 3700, pp. 393–401.
13. Bagavathiappan, S.; Lahiri, B.B.; Saravanan, T.; Philip, J.; Jayakumar, T. Infrared Thermography for Condition Monitoring—A Review. *Infrared Phys. Technol.* **2013**, *60*, 35–55. [CrossRef]
14. Asdrubali, F.; D’Alessandro, F.; Schiavoni, S. A Review of Unconventional Sustainable Building Insulation Materials. *Sustain. Mater. Technol.* **2015**, *4*, 1–17. [CrossRef]
15. Ibarra-Castanedo, C.; Sfarra, S.; Klein, M.; Maldague, X. Solar Loading Thermography: Time-Lapsed Thermographic Survey and Advanced Thermographic Signal Processing for the Inspection of Civil Engineering and Cultural Heritage Structures. *Infrared Phys. Technol.* **2017**, *82*, 56–74. [CrossRef]
16. Ohlsson, K.E.A.; Olofsson, T. Quantitative Infrared Thermography Imaging of the Density of Heat Flow Rate through a Building Element Surface. *Appl. Energy* **2014**, *134*, 499–505. [CrossRef]
17. Voogt, J.A.; Oke, T.R. Thermal Remote Sensing of Urban Climates. *Remote Sens. Environ.* **2003**, *86*, 370–384. [CrossRef]
18. Mirzaei, P.A.; Haghighat, F. Approaches to Study Urban Heat Island—Abilities and Limitations. *Build. Environ.* **2010**, *45*, 2192–2201. [CrossRef]
19. Grimmond, S. Urbanization and Global Environmental Change: Local Effects of Urban Warming. *Geogr. J.* **2007**, *173*, 83–88. [CrossRef]
20. Ardabili, S.; Abdolalizadeh, L.; Mako, C.; Torok, B.; Mosavi, A. Systematic Review of Deep Learning and Machine Learning for Building Energy. *Front. Energy Res.* **2022**, *10*, 786027. [CrossRef]
21. LeCun, Y.; Bengio, Y.; Hinton, G. Deep Learning. *Nature* **2015**, *521*, 436–444. [CrossRef]
22. Shlezinger, N.; Whang, J.; Eldar, Y.C.; Dimakis, A.G. Model-Based Deep Learning. *Proc. IEEE* **2023**, *111*, 465–499. [CrossRef]
23. De Bézenac, E.; Pajot, A.; Gallinari, P. Deep Learning for Physical Processes: Incorporating Prior Scientific Knowledge. *J. Stat. Mech. Theory Exp.* **2019**, *2019*, 124009. [CrossRef]
24. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet Classification with Deep Convolutional Neural Networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*. Available online: https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf (accessed on 19 August 2024). [CrossRef]
25. Luo, Q.; Gao, B.; Woo, W.L.; Yang, Y. Temporal and Spatial Deep Learning Network for Infrared Thermal Defect Detection. *Ndt E Int.* **2019**, *108*, 102164. [CrossRef]
26. Park, G.; Lee, M.; Jang, H.; Kim, C. Thermal Anomaly Detection in Walls via CNN-Based Segmentation. *Autom. Constr.* **2021**, *125*, 103627. [CrossRef]
27. Perez, H.; Tah, J.H.; Mosavi, A. Deep Learning for Detecting Building Defects Using Convolutional Neural Networks. *Sensors* **2019**, *19*, 3556. [CrossRef] [PubMed]
28. Tien, P.W.; Wei, S.; Darkwa, J.; Wood, C.; Calautit, J.K. Machine Learning and Deep Learning Methods for Enhancing Building Energy Efficiency and Indoor Environmental Quality—a Review. *Energy AI* **2022**, *10*, 100198. [CrossRef]
29. Wang, Z.; Hong, T.; Piette, M.A. Building Thermal Load Prediction through Shallow Machine Learning and Deep Learning. *Appl. Energy* **2020**, *263*, 114683. [CrossRef]
30. Somu, N.; Sriram, A.; Kowli, A.; Ramamritham, K. A Hybrid Deep Transfer Learning Strategy for Thermal Comfort Prediction in Buildings. *Build. Environ.* **2021**, *204*, 108133. [CrossRef]
31. Edalatifar, M.; Tavakoli, M.B.; Ghalambaz, M.; Setoudeh, F. Using Deep Learning to Learn Physics of Conduction Heat Transfer. *J. Therm. Anal. Calorim.* **2021**, *146*, 1435–1452. [CrossRef]
32. Gunay, B.; Shen, W.; Yang, C. Characterization of a Building’s Operation Using Automation Data: A Review and Case Study. *Build. Environ.* **2017**, *118*, 196–210. [CrossRef]
33. Ruiz, L.G.B.; Cuéllar, M.P.; Calvo-Flores, M.D.; Jiménez, M.D.C.P. An Application of Non-Linear Autoregressive Neural Networks to Predict Energy Consumption in Public Buildings. *Energies* **2016**, *9*, 684. [CrossRef]
34. Keramitsoglou, I.; Kiranoudis, C.T.; Ceriola, G.; Weng, Q.; Rajasekar, U. Identification and Analysis of Urban Surface Temperature Patterns in Greater Athens, Greece, Using MODIS Imagery. *Remote Sens. Environ.* **2011**, *115*, 3080–3090. [CrossRef]
35. Weng, Q.; Rajasekar, U.; Hu, X. Modeling Urban Heat Islands and Their Relationship with Impervious Surface and Vegetation Abundance by Using ASTER Images. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 4080–4089. [CrossRef]
36. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef]
37. Hersbach, H.; Bell, B.; Berrisford, P.; Hirahara, S.; Horányi, A.; Muñoz-Sabater, J.; Nicolas, J.; Peubey, C.; Radu, R.; Schepers, D. The ERA5 Global Reanalysis. *Q. J. R. Meteorol. Soc.* **2020**, *146*, 1999–2049. [CrossRef]
38. Li, J.; Hong, D.; Gao, L.; Yao, J.; Zheng, K.; Zhang, B.; Chanussot, J. Deep Learning in Multimodal Remote Sensing Data Fusion: A Comprehensive Review. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *112*, 102926. [CrossRef]
39. Wang, Y.; Long, M.; Wang, J.; Gao, Z.; Yu, P.S. Predrnn: Recurrent Neural Networks for Predictive Learning Using Spatiotemporal Lstms. *Adv. Neural Inf. Process. Syst.* **2017**, *30*. Available online: https://proceedings.neurips.cc/paper_files/paper/2017/file/e5f6ad6ce374177eef023bf5d0c018b6-Paper.pdf (accessed on 19 August 2024).

40. Wang, Y.; Wu, H.; Zhang, J.; Gao, Z.; Wang, J.; Philip, S.Y.; Long, M. Predrnn: A Recurrent Neural Network for Spatiotemporal Predictive Learning. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *45*, 2208–2225. [[CrossRef](#)]
41. Shi, X.; Chen, Z.; Wang, H.; Yeung, D.-Y.; Wong, W.-K.; Woo, W. Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting. *Adv. Neural Inf. Process. Syst.* **2015**, *28*. Available online: https://proceedings.neurips.cc/paper_files/paper/2015/file/07563a3fe3bbe7e3ba84431ad9d055af-Paper.pdf (accessed on 19 August 2024).
42. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2014**, arXiv:1412.6980.
43. Willmott, C.J.; Matsuura, K. Advantages of the Mean Absolute Error (MAE) over the Root Mean Square Error (RMSE) in Assessing Average Model Performance. *Clim. Res.* **2005**, *30*, 79–82. [[CrossRef](#)]
44. Chai, T.; Draxler, R.R. Root Mean Square Error (RMSE) or Mean Absolute Error (MAE)?—Arguments against Avoiding RMSE in the Literature. *Geosci. Model Dev.* **2014**, *7*, 1247–1250. [[CrossRef](#)]
45. Everingham, M.; Van Gool, L.; Williams, C.K.; Winn, J.; Zisserman, A. The Pascal Visual Object Classes (Voc) Challenge. *Int. J. Comput. Vis.* **2010**, *88*, 303–338. [[CrossRef](#)]
46. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [[CrossRef](#)]
47. Hore, A.; Ziou, D. Image Quality Metrics: PSNR vs. SSIM. In Proceedings of the 2010 20th International Conference on Pattern Recognition, Istanbul, Turkey, 23–26 August 2010; pp. 2366–2369.
48. Taylor, S.J.; Letham, B. Forecasting at Scale. *Am. Stat.* **2018**, *72*, 37–45. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.