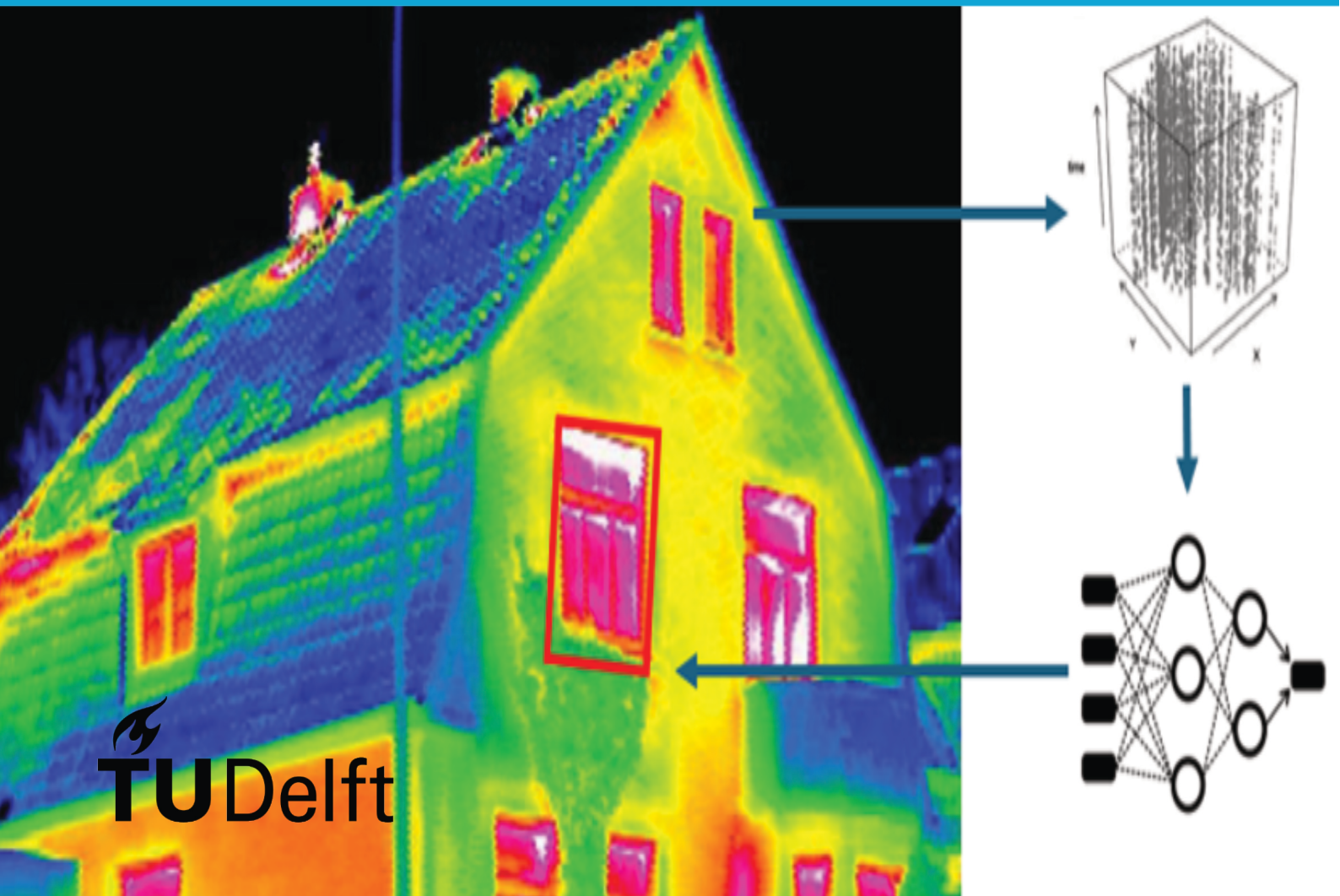
MSc thesis in Geomatics

Detecting building element / material through ground-based thermal imagery using Deep Neural Networks approach

Jiaoyang Wu

2026



MSc thesis in Geomatics

Detecting building element/material through ground-based thermal imagery using Deep Neural Networks approach

Jiaoyang Wu

June 2026

A thesis submitted to the Delft University of Technology in partial fulfillment of the requirements for the degree of Master of Science in Geomatics

Supervisors:	Dr. Azarakhsh Rafiee
	Dr. Regina Bokel
Co-reader:	PhD Amy Sterrenberg

Abstract

Building material identification plays an important role in applications such as sustainable building assessment, facade inspection, energy efficiency analysis, and urban environmental studies. Traditional material identification approaches mainly rely on optical imagery or geometric information, which are often limited in capturing the thermophysical characteristics of facade materials. Thermal imagery provides additional information related to heat transfer behavior, emissivity, and thermal inertia, offering an alternative perspective for building material analysis.

This thesis investigates the use of spatio-temporal thermal imagery combined with a Convolutional Long Short-Term Memory (Convolutional Long Short-Term Memory (ConvLSTM))-based deep learning framework for pixel-level building material classification. Ground-based thermal image sequences of building facades were collected under different environmental and temporal conditions using a thermal camera. Building materials including glass, concrete, and brick were manually annotated through Red Green Blue (RGB) imagery and Segment Anything Model (SAM)-assisted segmentation. Image registration techniques were subsequently applied to align the RGB and thermal image sequences, enabling the transfer of material labels into the thermal domain.

A ConvLSTM-based semantic segmentation model was developed to capture both spatial thermal distributions and temporal thermal dynamics from sequential thermal imagery. Different dataset splitting strategies and temporal datasets were evaluated to investigate the influence of spatial generalization and temporal variation on classification performance. The experimental results demonstrate that spatio-temporal thermal sequences provide meaningful discriminative information for building material classification. Under the random sequence-level split strategy, the proposed framework achieved a mean F1-score of approximately 0.686 and a mean Intersection over Union (IoU) of approximately 0.524 across the evaluated material classes.

The results indicate that temporal thermal behavior significantly contributes to distinguishing facade materials. Brick surfaces exhibited relatively stable and spatially uniform thermal characteristics, while glass surfaces demonstrated stronger temporal variation and environmental sensitivity. Concrete exhibited intermediate thermal behavior, making it comparatively more difficult to classify. The findings suggest that incorporating temporal thermal information improves segmentation performance compared with single-frame thermal analysis.

Acknowledgements

First of all, I would like to express my sincere gratitude to my supervisors, Dr. Azarakhsh Rafiee and Dr. Regina Bokel, for their continuous guidance, valuable suggestions, and supportive feedback throughout this thesis project. Their expertise and encouragement greatly helped me in completing this research.

I would also like to thank Amy Sterrenberg for kindly agreeing to serve as the co-reader of this thesis and for her time and support.

Special thanks are extended to my friends who provided valuable suggestions, discussions, and encouragement throughout the research and writing process.

Finally, I am indebted to my parents for their continuous support and encouragement.

Contents

List of Figures

List of Tables

Acronyms

RGB	Red Green Blue	v
ROI	Region of Interest	23
AI	Artificial Intelligence	8
ML	Machine Learning	8
ANN	Artificial Neural Network	9
CNN	Convolutional Neural Network	8
RNN	Recurrent Neural Network	13
LSTM	Long Short-Term Memory	8
ConvLSTM	Convolutional Long Short-Term Memory	v
SAM	Segment Anything Model	v
SIFT	Scale-Invariant Feature Transform	23
RANSAC	Random Sample Consensus	23
OA	Overall Accuracy	25
IoU	Intersection over Union	v
mIoU	Mean Intersection over Union	26
TP	True Positive	26
FP	False Positive	26
FN	False Negative	26
GPU	Graphics Processing Unit	24

1. Introduction

In recent years, advances in sensing technologies and deep learning have opened new possibilities for building envelope inspection and material analysis (Shariq and Hughes, 2020). Identifying and analyzing the materials used in construction plays a crucial role in promoting sustainability by facilitating material reuse, recycling, and reducing the environmental impact. Although traditional approaches to building element detection have relied heavily on optical images and laser scanning data, thermal images offer unique advantages (Wilson et al., 2023), such as the ability to detect hidden defects, assess insulation performance and analyze material properties based on heat signatures.

Infrared thermography has long been recognized as a powerful and non-destructive method for assessing energy loss, detecting moisture damage, and evaluating insulation performance in buildings (Balaras and Argiriou, 2002). However, its application in precise building material identification has been limited due to a lack of standardized methodologies and insufficient research on the thermal properties of different construction materials. The ability to analyze the spatiotemporal thermal dynamics of building elements offers a promising avenue to improve the assessment of energy efficiency and structural integrity (Seo et al., 2023).

This research aims to explore the suitability of a deep neural network approach, specifically using ConvLSTM models, to detect and classify building elements and materials through ground-based thermal imagery. Unlike static optical imagery, thermal imagery provides a rich dataset that captures the dynamic heat transfer characteristics of materials over time. By analyzing thermal variations at different times of the day, it becomes possible to distinguish materials such as brick, glass, and concrete based on their unique thermal retention and dissipation properties.

To achieve this, close-range thermal imagery will be collected from an existing building (e.g., the Architecture Faculty building) at high temporal resolution. A deep learning model, which incorporates both spatial and temporal features, will be trained to classify building materials based on their thermal signatures. This approach has the potential to improve automated building diagnostics, improve energy management strategies, and contribute to more sustainable building practices by optimizing material selection and life cycle assessment.

1.1. Motivation

As mentioned above, identifying and analyzing the elements/materials used in construction supports promoting sustainable practices in the building industry. This process aims to enhance the lifecycle of materials by facilitating their reuse, recycling, and repurposing, thereby minimizing waste and reducing the environmental impact. While optical imagery and laser scanning dataset have been widely applied for building element detection and

building analysis, the potential of thermal imagery for building material classification has not yet been fully explored. Thermal imagery provides additional information related to heat distribution and thermal behavior, which may improve the discrimination of materials with different thermal properties. Furthermore, the integration of sensing technologies and deep learning approaches, particularly spatio-temporal neural networks such as ConvLSTM, presents a promising opportunity to analyze thermal dynamics. This research is motivated by the potential to improve automated building material classification and contribute to more sustainable building assessment through the use of spatio-temporal thermal imagery and deep learning methods.

1.2. Objective and Scope

The primary objective of this thesis is to investigate the suitability of ConvLSTM-based deep learning models for detecting and classifying building elements using spatio-temporal thermal imagery. The research focuses on exploiting the thermal property of façade materials captured over time to improve material classification performance.

The scope of this research includes:

- Collecting and pre-processing a dataset of ground-based thermal image sequences together with manually labeled ground truth,
- Modeling the spatio-temporal thermal behavior of building materials using ConvLSTM networks,
- Extracting and analyzing thermal properties for building material classification,
- Focusing mainly on façade materials such as brick, concrete, and glass,
- Evaluating the performance of the ConvLSTM model under diurnal and seasonal thermal changes.

The aim of this thesis is to gain insight into the suitability of spatio-temporal thermal imagery and deep learning methods for automated building material classification. Furthermore, this research aims to contribute to more efficient building assessment and material analysis by providing a framework for thermal-based building material detection.

1.3. Research question

The main research question for this thesis is:

To what extent is a ConvLSTM-based deep learning model using spatio-temporal thermal imagery suitable for detecting and classifying different building materials and elements?

To comprehensively explore this question, several key factors influencing model performance must be considered, including data quality, network architecture, and temporal variations in thermal characteristics. Therefore, the following research subquestions are formulated:

- How can a representative and high-quality training dataset be collected for building material classification using thermal imagery?
- What ConvLSTM model architecture and hyperparameter settings yield the best classification performance for different building materials?
- How does the temporal variation in thermal imagery (e.g., daily and seasonal changes) impact the classification accuracy of different materials?
- What are the key thermal properties (e.g., specific heat, thermal inertia, spatial heat distribution) that contribute most to distinguishing between materials in thermal imagery?

1.4. Thesis outline

The thesis consists of 6 chapters. In Chapter 2 a literature review is provided. This chapter presents an overview of thermal image applications, building material identification, and deep learning architectures related to spatio-temporal image analysis. Chapter 3 details the methodology of this research. In Chapter 4, the technical implementation of the methodology is presented. In Chapter 5, the results of the different experiments are presented and analyzed. Chapter 6 discusses the key findings of this research, together with the limitations of the study and recommendations for future work.

2. Theoretical background and related work

2.1. Theoretical background

2.1.1. Infrared thermal imaging

Principles of infrared radiation

As shown in Figure 2.1, infrared radiation is electromagnetic radiation with wavelengths longer than visible light and shorter than microwaves. All objects with a temperature above absolute zero emit infrared radiation as a result of the thermal motion of atoms and molecules within the material (Mikaél'A, 2013). The intensity and spectral distribution of the emitted radiation are related to the temperature and physical properties of the object's surface. Infrared thermal imaging applies this principle to detect and visualize temperature distribution without direct contact.

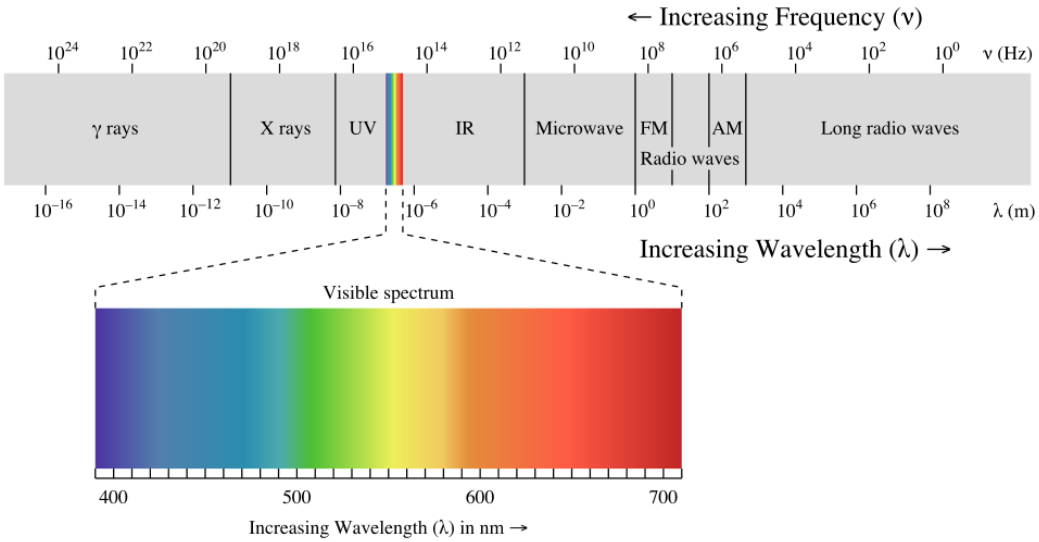


Figure 2.1.: Electromagnetic spectrum and visible light wavelength range

According to the Stefan–Boltzmann law (Wellons, 2007), the relationship between thermal radiation and surface temperature can be mathematically described in Equation 2.1.

$$E = \varepsilon \sigma T^4 \tag{2.1}$$

where:

- E is the emitted radiant energy,
- ε is the emissivity of the material surface,
- σ is the Stefan–Boltzmann constant, and
- T is the absolute temperature in Kelvin.

As shown in Equation 2.1, thermal radiation increases proportionally to the fourth power of temperature. Consequently, relatively small temperature variations can produce significant differences in emitted infrared radiation.

Emissivity varies among different materials and refers the effectiveness of a surface in emitting thermal radiation relative to an ideal blackbody. High-emissivity materials radiate thermal energy effectively, while low-emissivity and reflective surfaces reflect ambient thermal radiation. Therefore, thermal imagery is determined not only by temperature distribution but also by the thermophysical properties of surface materials.

Building materials such as concrete, brick, and glass possess distinct thermophysical properties (Pezeshki et al., 2018), including thermal conductivity, heat capacity, and thermal diffusivity. These properties affect how materials absorb, store, and release heat, resulting in different thermal patterns during temperature change. Such variations can potentially provide discriminative information for building material identification.

Infrared thermal cameras detect emitted infrared radiation and convert it into digital images representing temperature-related intensity values. Compared with RGB-based images, thermal imaging can present subsurface and thermophysical characteristics that are not directly observable in visible-light images, which makes infrared thermography particularly valuable (Chen et al., 2026) for building inspection and material analysis applications.

Thermal properties of building materials

As mentioned above, the thermal properties of materials, such as heat capacity, density, thermal conductivity, and thermal hysteresis behavior, influence the temperature distribution observed in infrared thermal images and can provide discriminative information for material identification.

For instance, specific heat capacity describes the amount of heat energy required to raise the temperature of a material by one degree Celsius (Wilhelm, 2010). Materials with high specific heat capacity generally heat up and cool down more slowly because they can store larger amounts of thermal energy. Density and thermal conductivity additionally influence the speed of heat transfer within the material.

As summarized in Table 2.1, several thermal properties of common building materials investigated in this research are compared.

Thermal Property	Brick	Concrete	Glass
Specific heat capacity (kJ/kgK)	0.841	0.879	0.792
Density	High	High	Relatively low
Heat transfer speed	Slow	Moderate	Fast
Daytime thermal response	Slow absorption	Moderate absorption	Fast increase
Nighttime thermal response	Gradual release	Moderate retention	Rapid cooling
Spatial heat distribution	Uniform	Uniform	Edge cooling
Thermal hysteresis	Strong	Moderate	Weak

Table 2.1.: Thermal properties of common building materials.

Brick exhibits relatively slow heat absorption and release due to its high thermal mass and significant thermal hysteresis. During daytime, the surface temperature increases gradually, while accumulated heat is slowly released during nighttime. This behavior produces relatively stable and spatially uniform thermal patterns.

Concrete demonstrates intermediate thermal behavior between brick and glass. Owing to its high density and moderate thermal conductivity, concrete can absorb and retain thermal energy more efficiently than glass while responding more rapidly than brick. As a result, concrete surfaces show moderate heating and cooling rates in thermal imagery.

Glass exhibits rapid thermal response characteristics because of its low specific heat capacity and high heat transfer efficiency. During daytime, the outer glass surface temperature rises rapidly, and thermal energy is quickly transferred through conduction. At night, the surface temperature decreases rapidly with limited heat retention. Furthermore, glass surfaces may exhibit non-uniform spatial temperature distributions, particularly near edges and window frames.

Research hypotheses

Based on the thermal properties discussed above, the following hypotheses are proposed in this study. These hypotheses form the theoretical foundation for the proposed ConvLSTM-based thermal image sequence analysis presented in this thesis.

- **Daily thermal variation:** Different building materials exhibit distinguishable heating and cooling patterns during day–night temperature cycles.
- **Thermal hysteresis:** Materials with high thermal mass exhibit stronger thermal hysteresis effects than low thermal mass materials.
- **Spatial thermal distribution:** Different building materials produce distinct spatial temperature distributions in thermal imagery.
- **Spatio-temporal discrimination:** Temporal thermal image sequences provide richer material-related information than single-frame thermal observations.

2.1.2. Building material identification

Building material identification technology aims to identify and classify different building materials based on their physical, spectral, geometric, or thermal properties. Traditional methods mainly rely on manual inspection and analysis (Liu et al., 2023), which are often time-consuming, labor-intensive, and difficult to apply on large spatial scales.

With the development of remote sensing and computer vision technologies, automated material identification methods are becoming increasingly popular. Image-based classification methods are among the most widely used because images contain rich visual and spatial information related to material properties. Optical images (e.g., RGB images) can capture surface texture, color, and reflectance patterns (Savelonas et al., 2022), while thermal infrared images provide information about thermophysical behavior, including differences in heat absorption, insulation, and emissivity.

Early image-based material classification methods typically used traditional machine learning algorithms, including Support Vector Machines (SVM) (Patel et al., 2019), Random Forests (RF) (Benedykciuk et al., 2021), and Decision Trees (Agarwal and Sharma, 2011). These methods often rely on manually designed features extracted from images, such as texture descriptors, edge information, color histograms, or spectral indices. While these methods can achieve good performance under controlled conditions, their effectiveness is often limited (Mehmood et al., 2022) when dealing with complex environments, varying lighting conditions, or materials with similar appearances.

In recent years, deep learning methods have demonstrated significant advantages in building material recognition tasks. Convolutional Neural Network (CNN) (Jogin et al., 2018) can automatically learn hierarchical spatial features directly from image data without manual feature engineering. Furthermore, recurrent architectures such as Long Short-Term Memory (LSTM) and ConvLSTM (Wang et al., 2022) can extract temporal patterns from sequential image data, making them suitable for analyzing changes in dynamic thermal behavior over time.

Depending on the classification objective, image-based material recognition can be performed at different levels, including image-level classification, object-level classification, and pixel-level classification. Among these methods, pixel-level annotation provides the most detailed representation because each pixel is assigned to a specific material category. This enables accurate material segmentation and spatial distribution analysis in building facades or scenes. Pixel-level classification is particularly important for thermal image analysis, as local temperature variations and spatial thermal patterns can provide crucial information for distinguishing different building materials.

2.1.3. Deep learning architecture

Machine learning

As a branch of Artificial Intelligence (AI), Machine Learning (ML) enables computer systems to learn patterns and relationships from data (Mehrotra, 2019). Traditional machine learning methods typically rely on hand-designed feature extraction, followed by classification algorithms such as Support Vector Machines (SVMs) and decision trees. However, hand-designed features may not effectively represent the complex spatial and temporal information in image datasets (Shanthamallu and Spanias, 2022).

As illustrated in Figure 2.2, machine learning is part of a broader hierarchy of artificial intelligence techniques, where deep learning and artificial neural networks represent increasingly specialized fields.

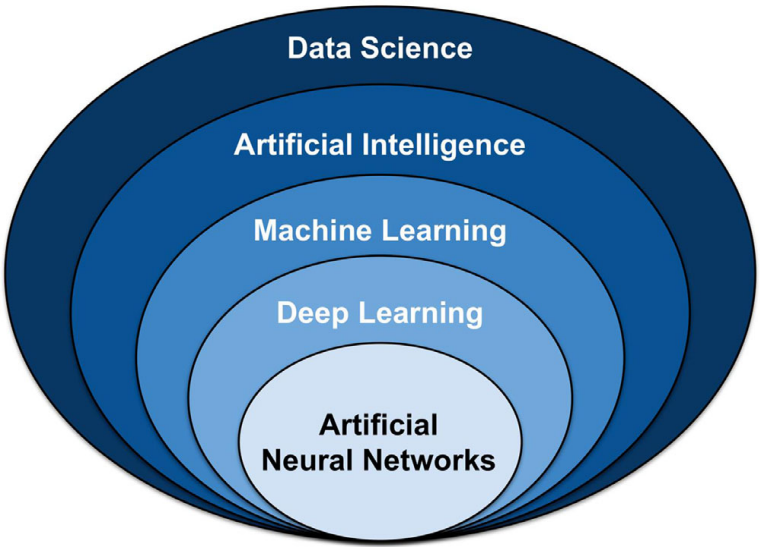


Figure 2.2.: Structure of select data science techniques (Choi et al., 2020).

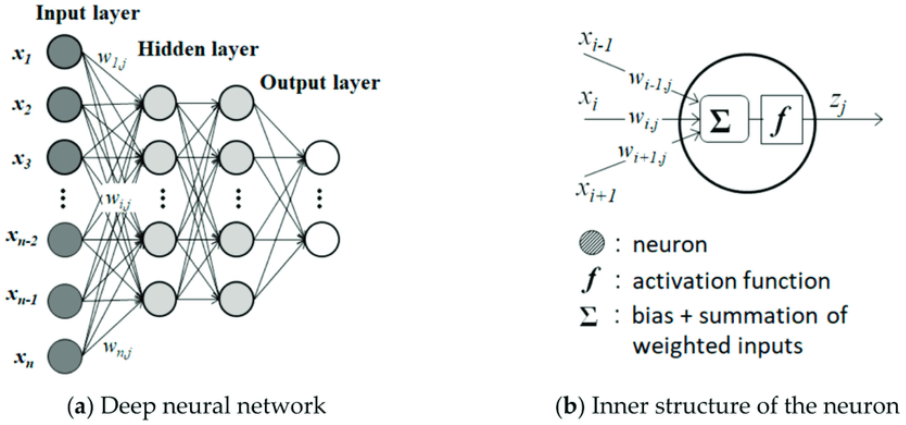
Deep learning

Deep learning is a subset of ML (Qamar and Zardari, 2023) that is based on Artificial Neural Networks (ANNs) with multiple hidden layers. Inspired by the structure of the human brain, deep learning models consist of interconnected artificial neurons organized into different network layers that process and transmit information. By stacking multiple hidden layers, deep learning models can learn hierarchical feature representations directly from raw input data (Schulz and Behnke, 2012). Lower network layers capture simple local patterns, whereas deeper layers learn more abstract and semantic representations.

Artificial Neural Network

In ANNs, data flows from the input layer to the output layer through one or more hidden layers (Krenker et al., 2011). Each layer consists of neurons that receive input, process it, and pass the output to the next layer. The layers work together to extract features, transform data, and make predictions.

The basic structure of ANNs and the perceptron model are illustrated in Figure 2.3. The perceptron represents the fundamental computational unit of an ANN, where multiple input signals are weighted, aggregated, and transformed through an activation function to produce the final output.



(**a**) Deep neural network (**b**) Inner structure of the neuron

Figure 2.3.: Structures of Artificial Neural Networks (a) and Perceptron (b). (Surendranathan, 2023)

The mathematical operation of a neuron can be described using the activation process shown in Equation 2.2. The output of a neuron y is obtained by applying an activation function $f(\cdot)$ to the aggregated input value a:

$$y = f(a) \tag{2.2}$$

where a represents the weighted sum of the input variables. The aggregated input is calculated in Equation 2.3 as:

$$a = \sum_{i=1}^{M} w_i x_i + w_0 \tag{2.3}$$

where x_i denotes the input variables, w_i represents the associated weights, and M is the total number of input features. The parameter w_0 is the bias term, which shifts the activation function and improves the flexibility of the network. The weights determine the strength of the connections between neurons and are iteratively optimized during the training.

During training, the network parameters are optimized by minimizing a loss function through backpropagation and gradient-based optimization algorithms. By iteratively updating network weights, ANNs gradually learn meaningful feature representations and improve prediction accuracy.

Convolutional Neural Network

One of the advantages of CNNs over traditional neural networks is their strong performance in processing structured visual data such as images, videos, and audio signals (Yao et al., 2018). CNNs are capable of automatically learning hierarchical feature representations from

input data (O'Shea and Nash, 2015), making them highly effective for computer vision applications. The general workflow and major components of CNNs are illustrated in Figure 2.4, including convolutional operations, feature extraction, pooling, and final classification through fully connected layers.

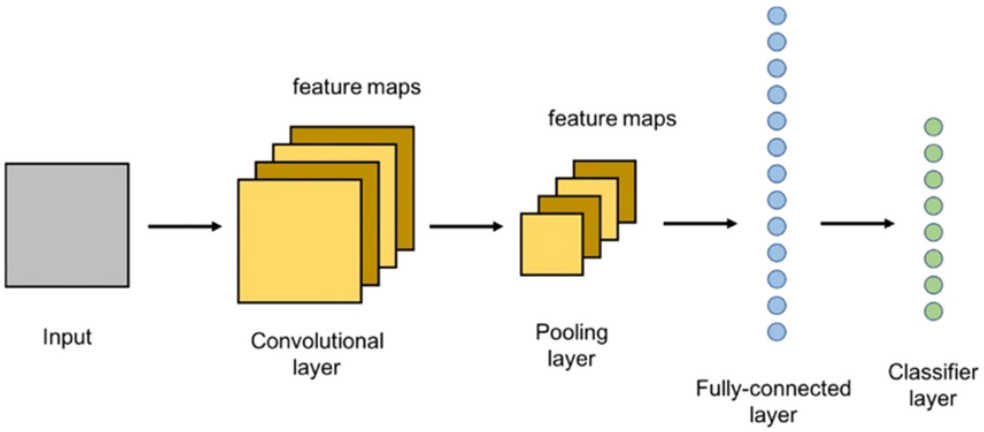


Figure 2.4.: The basic architecture of CNN. (Çekim et al., 2023)

The convolutional layer is the fundamental building block of a CNN and performs most of the computational operations within the network (Khan et al., 2020). It generally serves as the first layer of the architecture and may be followed by additional convolutional layers or pooling layers. As the network depth increases, CNNs progressively learn more complex and abstract features. Early layers usually capture low-level visual features such as edges, corners, and textures, while deeper layers identify larger structures and semantic patterns, eventually enabling object recognition.

The convolutional layer operates using three main components: the input data, convolution kernels (filters), and output feature maps. For example, a color image can be represented as a three-dimensional matrix consisting of image height, width, and color channels (RGB). A convolution kernel is a small matrix of trainable weights that slides across the image to detect specific local features. This process is referred to as convolution.

The convolution operation is illustrated in Figure 2.5, where a convolution kernel slides across the input image to extract local spatial features.

During convolution, the kernel performs element-wise multiplication with the corresponding region of the input image, followed by summation of the results. The generated values form an output feature map. The kernel then moves across the image according to a predefined stride until the entire image has been scanned. The convolution operation can be expressed as Equation 2.4:

$$S(i,j) = \sum_m \sum_n I(i-m, j-n) K(m,n) \tag{2.4}$$

where I denotes the input image, K represents the convolution kernel, and $S(i,j)$ is the resulting feature value at location (i,j). The weights of the convolution kernel remain shared while scanning the image, which significantly reduces the number of trainable parameters

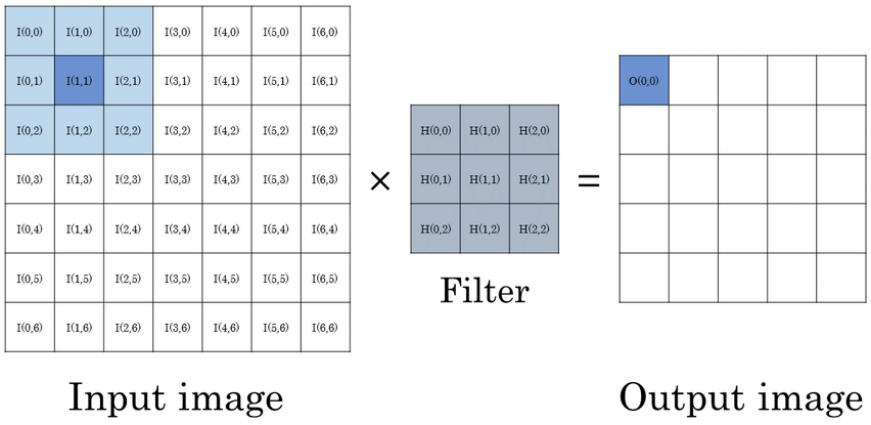


Figure 2.5.: Example of image convolution. (Baskin et al., 2017)

compared with fully connected networks. These kernel weights are optimized during training through backpropagation and gradient descent.

Several hyperparameters influence the behavior of convolutional layers (Tuba et al., 2021). The number of filters determines the depth of the output feature maps, where different filters learn different visual patterns. The stride controls the movement step size of the kernel across the input image, and larger stride values generally produce smaller output feature maps. Padding is another important parameter used to control the spatial size of the output. Zero-padding is commonly applied around image boundaries to preserve spatial dimensions during convolution. Common padding strategies include valid padding (no padding), same padding (output size equal to input size), and full padding (increased output size).

After each convolution operation, nonlinear activation functions are applied to introduce nonlinearity into the model. The Rectified Linear Unit (ReLU) is one of the most commonly used activation functions in CNNs because of its computational efficiency and ability to alleviate vanishing gradient problems.

CNNs often contain multiple convolutional layers stacked sequentially to form hierarchical feature representations. In deeper layers, the network can combine simple low-level features into more complex semantic structures. This hierarchical learning mechanism enables CNNs to effectively capture complex object structures.

Pooling layers, also known as downsampling layers, are commonly introduced between convolutional layers to reduce the spatial dimensions of feature maps. By decreasing the feature map size, pooling layers reduce computational cost and help prevent overfitting (Zafar et al., 2022). Unlike convolutional layers, pooling operations do not contain trainable weights. Instead, a pooling filter scans the input feature map and aggregates information within local regions.

An example of the max pooling operation is illustrated in Figure 2.6:

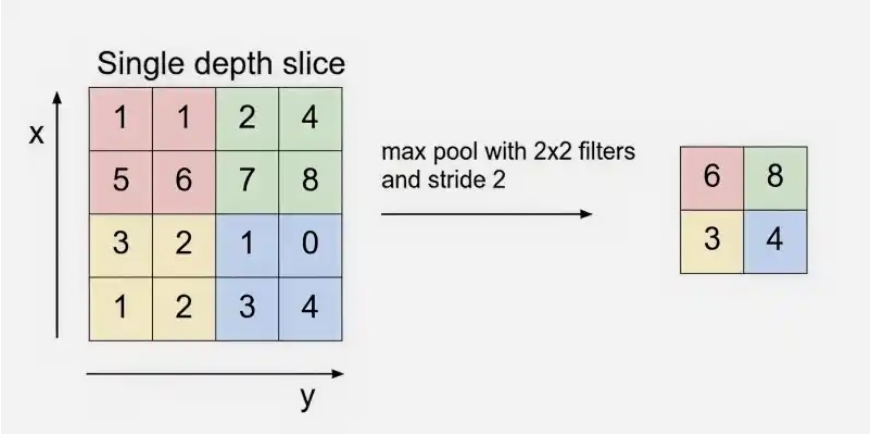


Figure 2.6.: Example of max pooling.

The fully connected layer is typically located at the end of the CNN architecture and performs the final classification task (Basha et al., 2020). In this layer, each neuron is fully connected to all neurons in the previous layer. The extracted high-level features from convolutional and pooling layers are flattened into a one-dimensional vector and passed into the fully connected layers for classification. Softmax activation is commonly used in the final output layer to convert the network outputs into probability distributions for different classes.

The structure of a fully connected layer is illustrated in Figure 2.7:

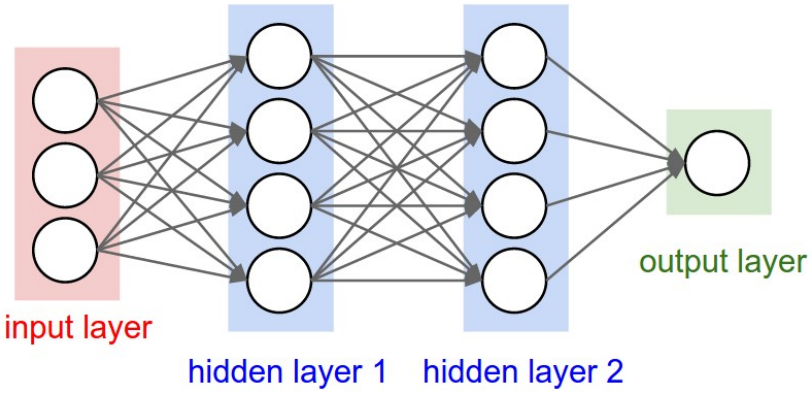


Figure 2.7.: Example of a fully connected network.

Recurrent Neural Network

Recurrent Neural Networks (RNNs) are a class of deep learning models specifically designed for processing sequential and time-series data (Zhou et al., 2019). Unlike traditional feedforward neural networks, which assume that input samples are independent from one another, RNNs are able to capture temporal dependencies between sequential inputs. The architecture and recurrent information flow of RNNs are illustrated in Figure 2.8:

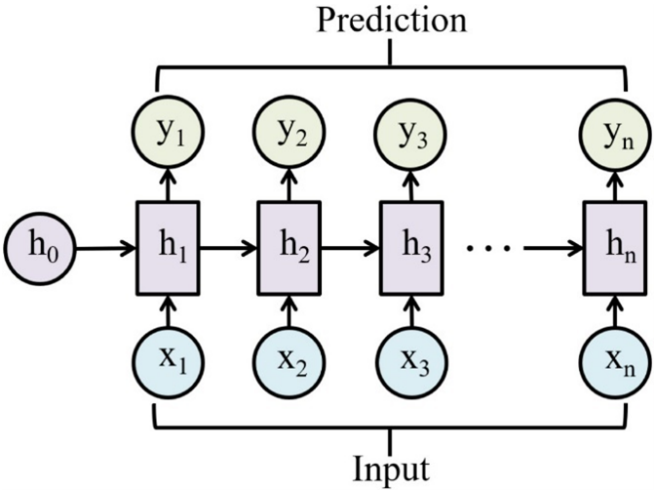


Figure 2.8.: The architecture of simple RNN. (Mienye and Jere, 2024)

In sequential data, the order of information is important because current outputs are often influenced by previous inputs. RNNs model this sequential relationship by preserving information from previous time steps (Fang et al., 2021) and using it to influence future predictions.

The key characteristic of an RNN is the hidden state, which acts as an internal memory of previous inputs (Ming et al., 2017). At each time step, the network processes the current input together with the hidden state from the previous step. This recurrent connection forms a feedback loop that enables the network to retain contextual information across the sequence. The hidden state update can be expressed in Equation 2.5:

$$h_t = f(W_{xh}x_t + W_{hh}h_{t-1} + b_h) \tag{2.5}$$

where x_t represents the current input at time step t, h_{t-1} denotes the hidden state from the previous time step, W_{xh} and W_{hh} are trainable weight matrices, b_h is the bias term, and f represents a nonlinear activation function such as *tanh* or ReLU. The output of the network can then be computed by Equation 2.6:

$$y_t = W_{hy}h_t + b_y \tag{2.6}$$

where y_t is the output at time step t, W_{hy} is the output weight matrix, and b_y is the output bias.

Another important property of RNNs is parameter sharing across all time steps. Unlike traditional feedforward neural networks, where each layer contains independent parameters, RNNs use the same weight matrices repeatedly throughout the sequence. This parameter sharing reduces the total number of trainable parameters and allows the network to generalize temporal patterns more effectively (Lai et al., 2018).

During training, RNNs utilize forward propagation and Backpropagation Through Time (BPTT) to optimize model parameters (Kag and Saligrama, 2021). BPTT is an extension of the traditional backpropagation algorithm. In this approach, the network computes errors at each time step and propagates them backward through the entire sequence. The accumulated gradients are then used to update the shared network parameters using optimization methods such as gradient descent.

Convolutional Long Short-Term Memory

ConvLSTM is an extension of the traditional LSTM network that is specifically designed for spatio-temporal data (Desai et al., 2022). In many real-world applications, data are collected continuously over time in the form of sequential images, including surveillance videos, satellite imagery, medical imaging, and environmental monitoring. These datasets contain both spatial information within individual frames and temporal dependencies between consecutive frames.

The internal memory structure and gating mechanism of LSTM are illustrated in Figure 2.9, where information is selectively retained, updated, and transferred through different gates across time steps.

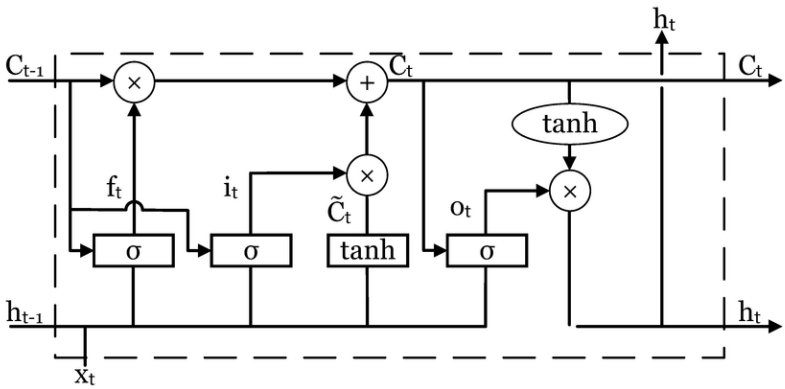


Figure 2.9.: LSTM cell structure. (Pisa et al., 2020)

Traditional LSTM networks are effective for modeling temporal dependencies in sequential data because they maintain hidden states that transfer information from previous time steps to subsequent ones. However, standard LSTMs process data as one-dimensional feature vectors and therefore cannot effectively preserve the spatial structure of image data. In contrast, CNNs are highly effective at extracting spatial features from images through convolution operations, but they do not explicitly model temporal relationships between consecutive frames.

ConvLSTM combines the advantages of CNNs and LSTMs (Yu et al., 2019) by replacing the matrix multiplication operations in conventional LSTM cells with convolution operations. As a result, spatial information is preserved while temporal dependencies are simultaneously learned. Instead of flattening images into one-dimensional vectors, ConvLSTM processes multidimensional image tensors directly, allowing the network to capture both spatial and temporal patterns within sequential image data.

The architecture of the ConvLSTM network is illustrated in Figure 2.10:

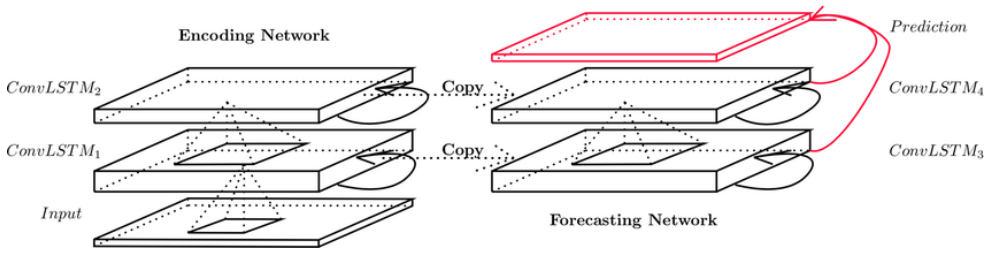


Figure 2.10.: A multi-layered encoding–forecasting ConvLSTM network. (Scheepens et al., 2023)

In a ConvLSTM cell, the input, hidden states, and memory states are represented as three-dimensional tensors (Ma et al., 2025). Convolution operations are applied during state transitions, enabling local spatial features to propagate through time. This structure makes ConvLSTM particularly suitable for tasks involving dynamic image sequences, such as video prediction, precipitation forecasting, traffic flow analysis, and thermal image sequence classification.

2.2. Related work

Thermal imaging has been widely used in the field of building inspection due to its non-invasive nature and ability to assess the thermal performance of building envelopes. Many studies have leveraged infrared thermography to detect heat loss (Moyano Campos et al., 2017), moisture infiltration (Rocha et al., 2018), and insulation deficiencies in buildings (Balaras and Argiriou, 2002). These applications primarily focus on thermal anomaly detection rather than material identification. However, the thermal behavior of building elements is inherently linked to their material properties, making thermal imagery a potentially valuable tool for material classification.

Traditional building material identification approaches have primarily relied on optical imagery. Optical imagery—often RGB photographs—allows for the extraction of surface-level visual features such as texture, color, and reflectance, which can be used to differentiate materials under well-lit and unobstructed conditions. Laser scanning (Dimitrov and Golparvar-Fard, 2014), particularly LiDAR, contributes high-resolution geometric information about surface structure and depth, supporting detailed modeling of building components. These approaches are often combined to improve accuracy; for example, Zahiri et al. utilized multispectral imagery together with LiDAR intensity data to characterize building materials by leveraging both spectral and geometric cues (Zahiri et al., 2021). While such methods have shown success in facade analysis and structural recognition, they are inherently limited to external appearance. They are sensitive to lighting, weather conditions, and occlusion, and offer little insight into the functional or thermophysical properties of materials. Furthermore, materials with similar visual characteristics (e.g., painted concrete vs. plaster) may be indistinguishable through optical or geometric data alone (Rashidi et al., 2016). These limitations highlight the potential of thermal imaging, which captures the dynamic heat

transfer behavior of materials and can reveal subsurface or performance-related differences not observable through traditional modalities.

With the rise of deep learning, CNNs have demonstrated remarkable success in image classification tasks, including those involving infrared data. In the context of building analysis, CNNs have been applied for automatic damage detection (Seo et al., 2023), heat leak localization (Waqas and and, 2024), and even building facade segmentation (Wang et al., 2024). However, these models typically work with static images, which do not capture the temporal dynamics of thermal behavior.

To address this limitation, researchers have turned to RNNs, which are specifically designed to model sequential datasets. RNNs and their advanced variants, such as LSTM networks, are well-suited for capturing temporal dependencies in time-series data (Junliang, 2022). LSTM has been widely applied in domains such as speech recognition (Saon et al., 2021), machine translation, and sensor data modeling (Moustapha and Selmic, 2008).

Building on this foundation, ConvLSTM extends LSTM by incorporating convolutional operations within its gating mechanisms, making it suitable for spatio-temporal image sequence modeling (Moustapha and Selmic, 2008). ConvLSTM has been applied to a range of tasks beyond video analysis—for instance, Wang et al. employed ConvLSTM for polarimetric synthetic aperture radar (PolSAR) image classification by feeding sequences of polarization coherent matrices in the rotation domain into a ConvLSTM-based network (Wang et al., 2018). Similarly, Hu et al. proposed deep ConvLSTM architectures for hyperspectral image (HSI) classification, effectively extracting spatial–spectral features through 2D and 3D variants to model long-range spectral dependencies while preserving spatial structure (Hu et al., 2020). However, its application in thermal image-based building material classification remains largely unexplored.

In summary, while previous work has made progress in using thermal imaging, CNNs, and RNN-based models independently, few studies have integrated spatio-temporal modeling of thermal data using ConvLSTM for the purpose of material classification. This research aims to bridge this gap by employing ConvLSTM networks to detect and classify building materials based on their thermal dynamics.

3. Methodology

This chapter presents the methodology adopted for building material identification using ground-based thermal image sequences and deep learning techniques. An overview of the overall research workflow is illustrated in Figure 3.1. The purpose and rationale of each step are introduced in this chapter, while the detailed implementation procedures and experimental settings are described in Chapter 4.

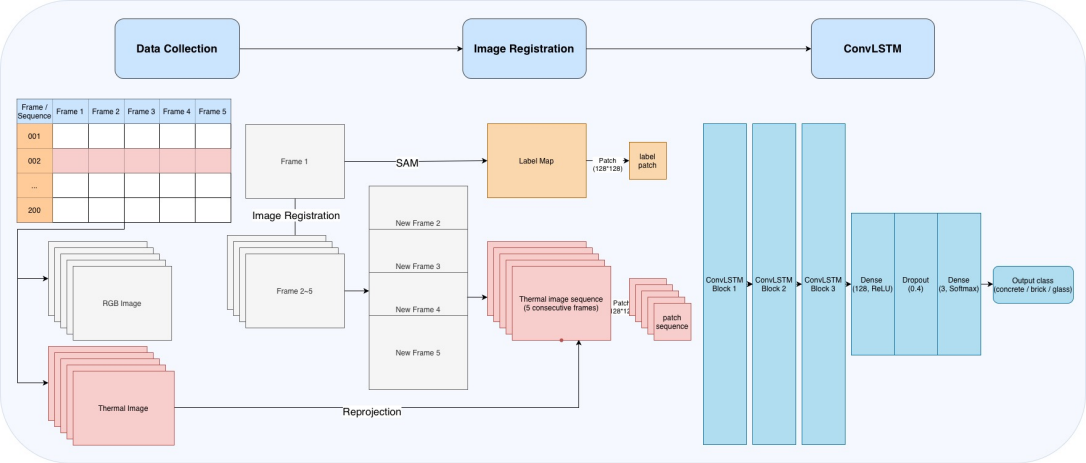


Figure 3.1.: Workflow.

The methodology of this study can be divided into four parts: (1) data collection, (2) data pre-processing, (3) ConvLSTM modeling, and (4) evaluation of classification results. The first part focuses on collecting raw thermal imaging and RGB image sequences using the thermal imaging camera. Thermal data were collected from building facades containing different building materials and covering various time and weather conditions. Subsequently, the collected data underwent pre-processing procedures, including image registration, dataset organization, and training label generation. After pre-processing, a ConvLSTM-based model was developed to learn the spatio-temporal thermal characteristics of building materials from sequential thermal imagery. Finally, the performance of the proposed model was evaluated using multiple classification metrics to assess the effectiveness of thermal temporal information for building material identification.

3.1. Data collection

3.1.1. Thermal imaging camera

The thermal dataset used in this study was collected using a FLIR E96 handheld thermal imager, as shown in Figure 3.2. This imager provides a raw thermal resolution of 640 × 480 pixels and supports radiometric thermal recording, capable of storing the temperature of each pixel in the captured thermal image. The FLIR E96 also provides high thermal sensitivity and interchangeable lens options, making it suitable for close-range building facade inspection and thermal pattern analysis. More detailed specifications are available on the official FLIR product page: https://www.flir.com/en-eu/products/e96/.

Figure 3.2.: FLIR E96 thermal imager.

3.1.2. Photo shooting plan

The thermal image dataset used in this research were collected from nine different buildings located across the campus Delft University of Technology (TU Delft) in Delft, the Netherlands from February to April 2026. The data acquisition locations included the Bouwkunde (Faculty of Architecture and the Built Environment), Pulse, IDE (Faculty of Industrial Design Engineering), CEG (Faculty of Civil Engineering and Geosciences), Echo, TNW (Faculty of Applied Sciences), Building 26, Korvezeestraat, and Building 35.

Figure 3.3 illustrates the nine acquisition locations used for thermal image collection in this research.

(a) BK (b) Pulse (c) IDE

(d) CEG (e) Echo (f) TNW

(g) Building 26 (h) Korvezeestraat (i) Building 35

Figure 3.3.: Acquisition locations in TU Delft Campus.

Figure 3.4 illustrates the Bouwkunde (Faculty of Architecture and the Built Environment) using Bouwkoude 3D Tiles (LoD1.2) (Peters et al., 2022) together with Google Maps orthophoto data. Among all acquisition locations, BK served as the primary data collection site and contained the largest number of captured thermal image sequences because of its diverse facade materials and complex architectural structures.

In order to capture temporal thermal variations, thermal image sequences were collected over multiple dates under different weather and temperature conditions. Approximately 200 RGB–thermal image pairs were collected during each shooting, resulting in a total dataset of 1000 image samples.

In addition, the shooting time and weather conditions were recorded for each frame so that reliable solar position information could be calculated using the NOAA Solar Position Calculator. A tripod was used during data collection to maintain a stable camera height and viewing angle for repeated observations. This setup reduced geometric inconsistencies between temporal image sequences and improved the quality of subsequent RGB–thermal image registration.

Table 3.1 summarizes the data collection time and weather conditions during the shooting

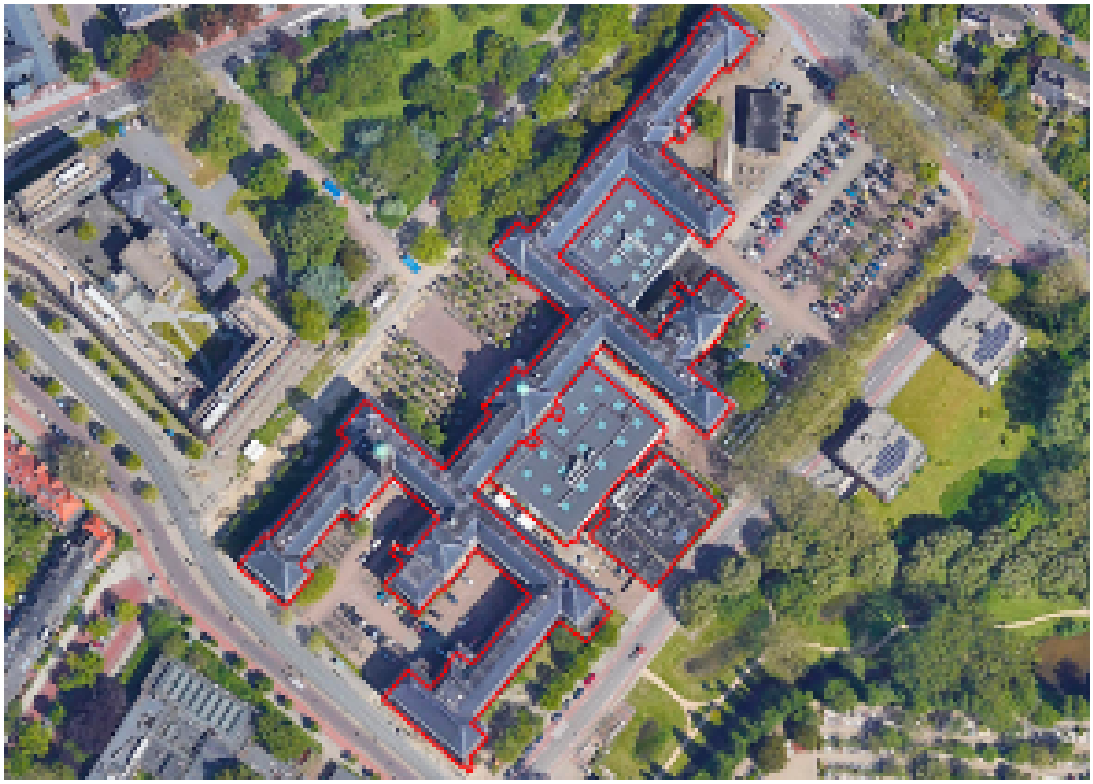


Figure 3.4.: Visualization of Bouwkunde (Faculty of Architecture and the Built Environment).

time. Changes in weather conditions can reveal different thermal responses of building materials and help to analyze their spatiotemporal thermal characteristics.

Table 3.1.: Data collection schedule.

Date	Time	Weather / Temperature	Images
11 Feb.	09:00–15:00	Moderate drizzle / 2–3°C	200
18 Feb.	09:00–15:00	Overcast / −2–3°C	200
25 Feb.	09:00–15:00	Overcast / 8–10°C	200
4 Mar.	09:00–14:00	Mist / 7–12°C	200
11 Mar.	10:00–14:00	Cloudy / 7°C	200

3.2. Data pre-processing

3.2.1. Dataset structure

The collected dataset was organized into sequence-based folders in order to preserve the temporal relationships between thermal images. Each image group contains synchronized RGB and thermal image pairs captured from the same place. The RGB images were mainly

used for visualization and material labeling, while the thermal images were used as the primary input for ConvLSTM modeling.

The dataset consists of five acquisition batches, namely data_1 to data_5, collected from multiple locations within the TU Delft campus. Each image group contains one RGB image and one corresponding thermal image. The detailed structure of the image dataset is summarized in Table 3.2.

Table 3.2.: Structure of the image dataset.

Batch	Location	Image groups	Images
data_1–data_5	BK	40 per batch	80 per batch
	Pulse	10 per batch	20 per batch
	IDE	25 per batch	50 per batch
	CEG	30 per batch	60 per batch
	Echo	30 per batch	60 per batch
	TNW	20 per batch	40 per batch
	Building 26	20 per batch	40 per batch
	Kor	15 per batch	30 per batch
	Building 35	10 per batch	20 per batch
Total	9 locations	**200 groups per batch**	**400 images per batch**

3.2.2. Image registration

Since the thermal image sequences were manually captured using a handheld thermal imaging camera, slight differences in camera position, viewing angle, and image scale occurred between consecutive acquisitions. Therefore, image registration was required to spatially align the RGB and thermal image pairs and reduce geometric inconsistencies between temporal image sequences.

In this research, the image sequence collected in data_1 was used as the reference dataset, while the image sequences from data_2 to data_5 were geometrically aligned to the corresponding reference images. Region of Interest (ROI)-based image registration was performed for each image sequence. The RGB image was used as the reference image, while the thermal image was geometrically transformed to align with the RGB image. Feature-based image registration methods, including Scale-Invariant Feature Transform (SIFT) and Random Sample Consensus (RANSAC), were employed to establish corresponding feature points between image pairs.

After feature matching, homography transformation matrices were computed to estimate the geometric relationship between the RGB and thermal images. The calculated transformations were subsequently applied to the thermal images to achieve pixel-level alignment.

3.3. Training data labeling

Pixel-level material labels were generated for the thermal image sequences in order to train the semantic segmentation model. In this research, the target classes included glass, concrete, brick, and background regions. Since thermal images often contain limited texture

information and unclear material boundaries, RGB images were primarily used during the labeling process to improve labeling accuracy.

The labeling workflow was performed on the registered RGB image sequences obtained from the image registration step. For each sequence, the first aligned RGB image was selected as the reference image for annotation. The generated labels were subsequently transferred to the corresponding thermal images due to the established pixel-level alignment between RGB and thermal image pairs.

SAM is a foundation segmentation model proposed by Meta AI for general-purpose image segmentation. SAM is capable of generating high-quality object masks from different types of prompts, including points, bounding boxes, and masks (Kirillov et al., 2023). In this research, SAM was used as an interactive labeling tool to assist the pixel-level annotation of building materials. By manually selecting positive and negative prompt points on RGB images, SAM automatically generated segmentation masks for facade materials such as glass, concrete, and brick. The generated masks were subsequently refined manually to improve labeling accuracy and consistency.

The final label maps were stored as pixel-level classification masks, where background pixels were assigned label 0, glass pixels label 1, concrete pixels label 2, and brick pixels label 3. The labeled thermal image sequences were subsequently used for ConvLSTM model training and evaluation.

3.4. ConvLSTM modeling

3.4.1. Model building

The proposed model was designed for pixel-level semantic segmentation of thermal image sequences. The input data consist of sequential thermal image patches with a sequence length of five frames. Each input sample was represented as a tensor with dimensions (B, T, C, H, W), where B denotes the batch size, T denotes the temporal sequence length, C denotes the number of image channels, and H and W represent the spatial dimensions of the image patch.

The ConvLSTM architecture used in this research consists of a ConvLSTM encoder followed by convolutional prediction layers. The ConvLSTM cell replaces the fully connected operations in traditional LSTM networks with convolutional operations, enabling simultaneous extraction of spatial and temporal features from thermal image sequences. The hidden states were updated iteratively across consecutive thermal frames to model temporal thermal variations of building materials.

After feature extraction, convolutional prediction layers were used to generate pixel-level material classification maps. The final output layer predicts four semantic classes, including background, glass, concrete, and brick. ReLU activation functions were applied in intermediate convolutional layers to introduce nonlinear feature learning.

During training, thermal image patches were extracted from the registered thermal image sequences. Patch-based training was adopted to reduce Graphics Processing Unit (GPU) memory consumption and increase the number of training samples. In addition, background-only patches were partially removed in order to reduce class imbalance problems during training.

To address the severe class imbalance between background regions and building material classes, a masked focal loss function was employed during model training. In this approach, background pixels were excluded from the loss calculation, while focal loss weighting was used to emphasize difficult training samples and underrepresented material classes.

3.4.2. Parameter optimization

In this research, Optuna was employed for automated hyperparameter optimization through multiple experimental trials (Akiba et al., 2019). The hyperparameter search space used for parameter optimization is summarized in Table 3.3.

Table 3.3.: Hyperparameter search space for ConvLSTM optimization.

Parameter	Search space	Type
Learning rate	$1 \times 10^{-5} - 5 \times 10^{-3}$	Log-uniform
Weight decay	$1 \times 10^{-6} - 1 \times 10^{-2}$	Log-uniform
Patch size	$\{64, 128\}$	Categorical
Loss function	{CE, Dice, CE+Dice, Weighted CE, Focal}	Categorical
Epochs	20	Fixed
Batch size	4	Fixed
Sequence length	5	Fixed
Background threshold	0.99	Fixed

During hyperparameter optimization, the number of training epochs was fixed to 20 for each trial in order to reduce computational cost while maintaining consistent comparisons between experiments. The Adam optimizer was employed for model optimization.

In addition, two different dataset splitting strategies were investigated. The first approach employed random sequence splitting, where all image sequences were randomly divided into training, validation, and test datasets using a ratio of 70%, 20%, and 10%, respectively.

The second approach employed group-based sequence splitting according to the spatial acquisition locations. In this strategy, image sequences from each building group were independently divided into training, validation, and test subsets while preserving the spatial grouping structure of the dataset. This approach reduces the risk of spatial information leakage between datasets and provides a stricter evaluation of model generalization across different facade locations.

3.5. Evaluation

The performance of the proposed model was evaluated by comparing the predicted material classification maps with the corresponding ground-truth labels from the test dataset. In this research, building material classification was formulated as a pixel-level semantic segmentation task. Therefore, the evaluation of the proposed model was performed on a per-pixel basis. The following metrics are observed to evaluate the performance of the model.

Overall Accuracy (OA) represents the proportion of correctly classified pixels relative to the total of pixels. It is calculated as shown in Equation 3.1 (Gonzalez et al., 2020).

$$OA = \frac{\sum_{i=1}^{N} TP_i}{\text{Total of pixels}} \tag{3.1}$$

where N represents the total number of classes and TP_i denotes the number of correctly classified pixels for class i.

However, OA alone may provide misleading results in multi-class semantic segmentation tasks when the dataset is imbalanced. In this research, background pixels occupy a large proportion of the thermal images, while some building material classes contain fewer pixels. Therefore, additional class-wise evaluation metrics were considered to better assess the classification performance for individual materials.

Recall represents the proportion of correctly classified pixels relative to the total number of ground-truth pixels belonging to a given class. Recall is calculated as shown in Equation 3.2 (Rashidi et al., 2016).

$$R = \frac{TP}{TP + FN} \tag{3.2}$$

where True Positive (TP) denotes true positives and False Negative (FN) denotes false negatives.

Precision represents the proportion of correctly predicted pixels among all predicted pixels for a given class. Precision is defined as shown in Equation 3.3 (Rashidi et al., 2016).

$$P = \frac{TP}{TP + FP} \tag{3.3}$$

where False Positive (FP) denotes false positives.

Based on precision and recall, the F1-score was additionally computed to provide a balanced evaluation metric for each material class. The F1-score combines precision and recall into a single value and is calculated as shown in Equation 3.4 (Yan et al., 2024).

$$F1 = \frac{2 \times P \times R}{P + R} \tag{3.4}$$

In addition, IoU was also used to evaluate the overlap between the predicted segmentation regions and the corresponding ground-truth labels. IoU is calculated as shown in Equation 3.5 (Zhang et al., 2022).

$$IoU = \frac{TP}{TP + FP + FN} \tag{3.5}$$

A larger IoU value indicates better segmentation quality. The Mean Intersection over Union (mIoU) was obtained by averaging the IoU values across all classes.

Furthermore, confusion matrices were generated to analyze the classification performance for different material categories, including brick, concrete, and glass (Perez et al., 2019). The confusion matrix provides a detailed visualization of class-wise prediction errors and helps identify material classes that are difficult to distinguish due to similar thermal characteristics.

Since background pixels occupy a large proportion of the images, additional evaluations were also conducted by excluding background regions from the metric calculation. This approach provides a more representative assessment of the classification performance for the target building materials.

4. Implementation

4.1. Training data preparation

Data Collection

As described in Section 3.1.2, raw thermal image data were collected using the FLIR E96 handheld thermal imager from multiple buildings located across the TU Delft campus. The complete dataset consists of 200 image sequences collected during five different acquisition sessions under varying weather and environmental conditions.

Figure 4.1 illustrates an example of the collected RGB and thermal image pairs.

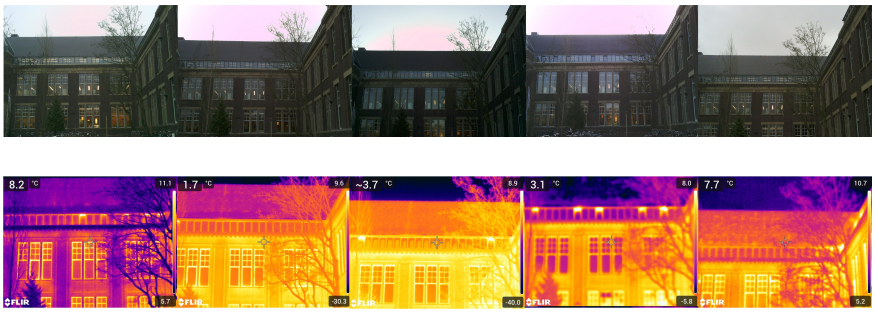


Figure 4.1.: Examples of Sequence 001.

Image Registration

As described in Section 3.2.2, image registration was performed to reduce geometric inconsistencies between repeated frames after data collection. The image sequences collected in data_1 were used as the reference dataset, while the corresponding image sequences from data_2 to data_5 were geometrically aligned to the reference images.

Figure 4.2 illustrates an example of the image registration process.

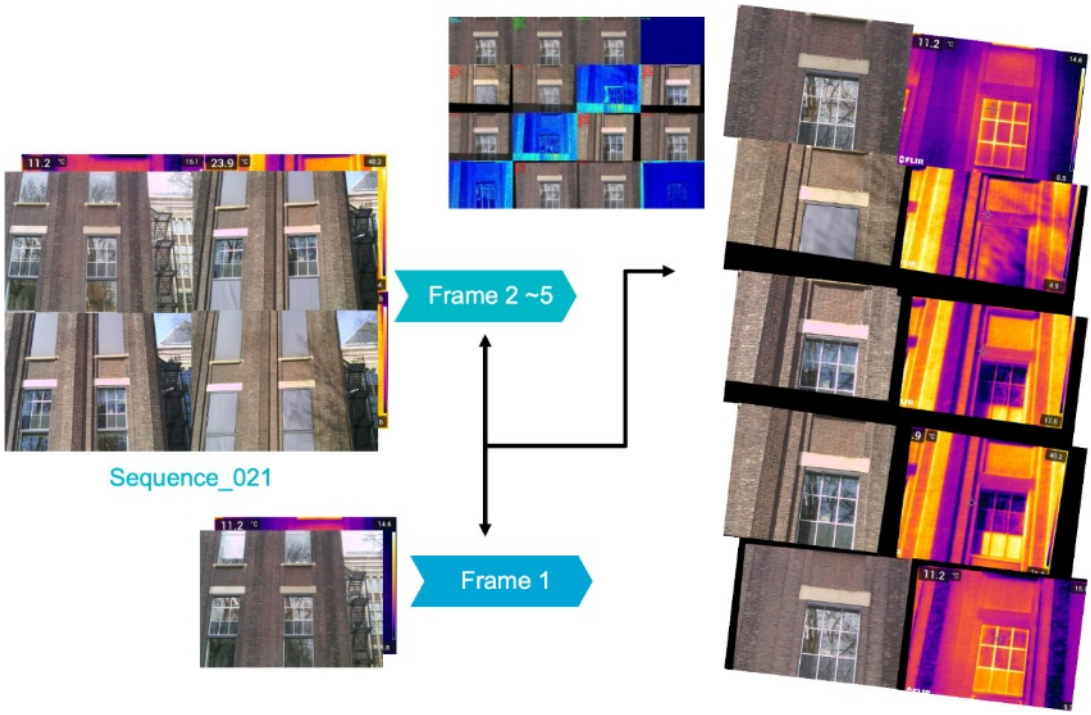


Figure 4.2.: The image registration process of Sequence 021.

SAM Labeling

As described in Section 3.3, pixel-level material labeling was performed using SAM. Since thermal images contain limited texture information and unclear material boundaries, the labeling process was conducted primarily on the aligned RGB images. The generated labels were transferred to the corresponding thermal images through the established image registration relationship.

In this research, the pretrained SAM model `sam_vit_b_01ec64.pth` was employed for interactive segmentation. Positive and negative prompt points were manually selected on the RGB images to guide the segmentation process, while SAM automatically generated the corresponding segmentation masks for facade materials such as glass, concrete, and brick.

Figure 4.3 illustrates examples of the SAM-assisted labeling workflow and the generated segmentation masks.

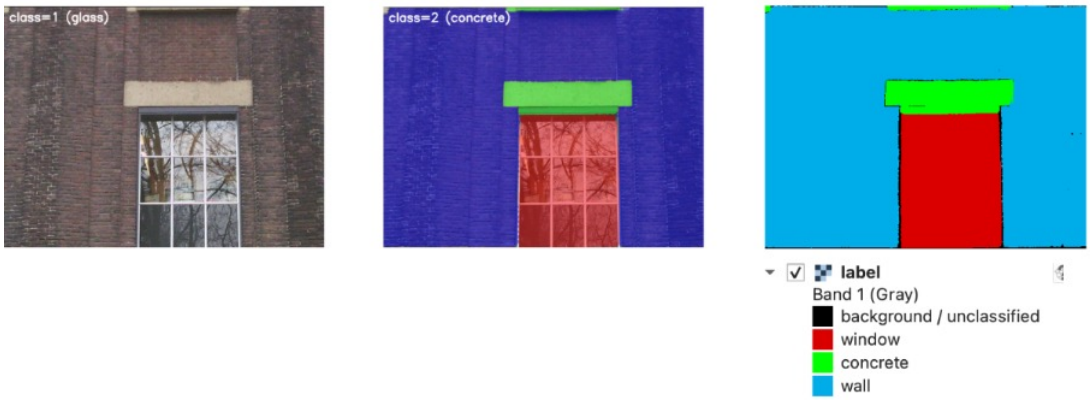


Figure 4.3.: Examples of SAM-based material labeling.

4.2. Hyperparameter setting

According to the hyperparameter optimization process described in Section 3.4.2, Optuna was employed to search for the optimal parameter configuration for the proposed ConvLSTM model. Multiple experimental trials were conducted using different combinations of learning rates, patch sizes, loss functions, and regularization parameters. The detailed hyperparameter search space was summarized in Table 3.3.

Two different dataset splitting strategies were additionally investigated during training, including random sequence splitting and group-based sequence splitting. The random split strategy achieved better overall segmentation performance and was therefore selected as the final training configuration.

After hyperparameter optimization, the best-performing parameter configuration obtained for the dataset is summarized in Table 4.1.

Table 4.1.: Final hyperparameter configuration for ConvLSTM training.

Parameter	**Value**
Experiment strategy	Random split
Learning rate	0.001
Weight decay	4×10^{-5}
Patch size	64
Stride	64
Loss function	Masked focal loss
Epochs	70
Background threshold	0.99
Number of classes	4
Ignore background in loss	True

In the final training configuration, masked focal loss was employed to address the severe class imbalance problem caused by the large proportion of background pixels in thermal

images. Background pixels were excluded from the loss calculation during training, while focal loss weighting emphasized difficult samples and underrepresented material classes.

Furthermore, a patch size of 64×64 together with a stride size of 64 was selected to balance computational efficiency and spatial feature preservation. The optimized parameter configuration was used for the final ConvLSTM training and evaluation experiments.

4.3. Python libraries and software tools

This research primarily utilizes Python, along with several open-source libraries and software tools, for image processing, deep learning, data visualization, and hyperparameter optimization. The overall workflow includes thermal image preprocessing, image registration, SAM-assisted labeling, ConvLSTM model training, and evaluation.

The Python libraries and software tools used in this research are summarized in Table 4.2.

Table 4.2.: Python libraries and software tools used in this research.

Tool / Library	**Application**
Python	Overall implementation
PyTorch	ConvLSTM model training
OpenCV	Image registration and preprocessing
NumPy	Array and tensor processing
Matplotlib	Result visualization
Optuna	Hyperparameter optimization and search
SAM	Interactive material labeling
QGIS and Google Maps	Acquisition location visualization and management

5. Results

5.1. Full dataset evaluation

Table 5.1.: Evaluation metrics for the two dataset splitting strategies.

Split	Class	Precision	Recall	F1-score	IoU
Random	Glass	0.712	0.726	0.719	0.561
	Concrete	0.603	0.672	0.636	0.467
	Brick	0.745	0.668	0.704	0.543
Group-based	Glass	0.630	0.766	0.691	0.528
	Concrete	0.639	0.498	0.560	0.389
	Brick	0.592	0.678	0.632	0.461

The evaluation results presented in Table 5.1 demonstrate noticeable performance differences between the random sequence splitting strategy and the group-based sequence splitting strategy.

Overall, the random segmentation strategy achieved better classification performance across most evaluation metrics. In particular, it produced higher F1-scores and IoU values in all three building material categories, indicating more stable segmentation performance and greater material discrimination. One possible explanation is that, under the random segmentation strategy, image sequences of similar spatial locations can appear simultaneously in both the training and test sets. Therefore, the thermal characteristics of the training and test samples remain relatively similar, facilitating model generalization.

Among all target materials, brick achieved the best overall segmentation performance under both strategies. Under the random segmentation strategy, brick achieved a precision of 0.745, a recall of 0.668, and f1-score of 0.704. This result indicates that brick surfaces have relatively stable and distinguishable thermal patterns compared to other facade materials. The higher IoU value also suggests good spatial overlap between the predicted and actual brick regions.

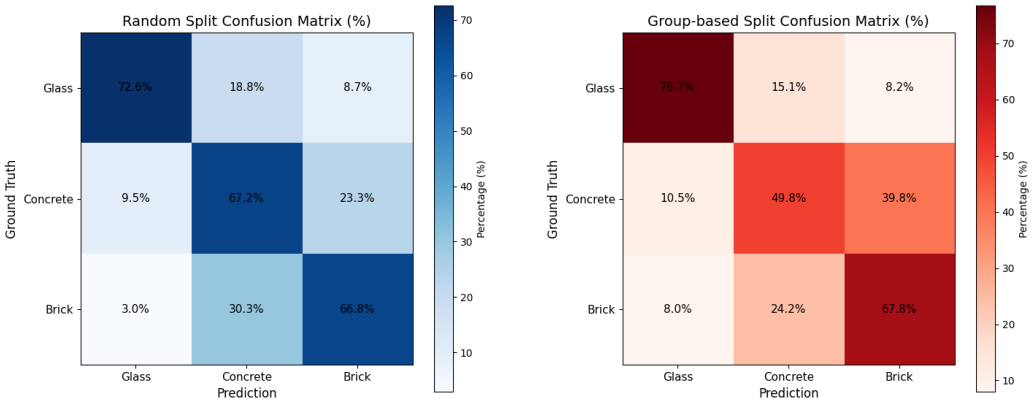
Concrete achieved moderate classification performance in both experimental environments. While concrete had relatively high recall, its precision was relatively low. This indicates that pixels from many other material categories were incorrectly classified as concrete. One possible explanation is that the thermal behavior of concrete is similar to that of other exterior cladding materials, especially under varying environmental conditions.

Among the three material categories, glass exhibited the most inconsistent classification performance. With the random segmentation strategy, glass showed relatively high precision and recall. However, its performance decreased significantly with the group-based segmentation strategy. This suggests that the thermal characteristics of glass are highly sensitive to

environmental factors such as reflection, sunlight, and viewing angle, which vary considerably between different buildings.

Overall, the comparative results demonstrate that building material classification based on thermal images remains strongly influenced by spatial and environmental variations. Although the proposed ConvLSTM model successfully learned significant spatio-temporal thermal features, the results also show that improving spatial generalization capability remains a major challenge for semantic segmentation based on thermal images.

Figure 5.1 shows the normalized confusion matrices for the random sequence splitting strategy and the group-based sequence splitting strategy, respectively.



(a) Random split

(b) Group-based split

Figure 5.1.: Normalized confusion matrices for the random sequence splitting strategy and the group-based sequence splitting strategy.

The random split strategy achieved better overall classification consistency across all material classes. In particular, brick achieved the highest classification accuracy, while concrete exhibited moderate confusion with brick regions. Glass achieved relatively stable classification performance under the random split strategy.

In contrast, the group-based split strategy produced noticeably larger confusion between concrete and brick materials. This observation indicates that the thermal characteristics learned from specific buildings may not generalize effectively to unseen facade locations.

5.2. Difference of training dataset

Figure 5.2 illustrates the training and validation loss curves for both the random sequence splitting strategy and the group-based sequence splitting strategy.

As shown in Figure 5.2, both training strategies demonstrate continuous decreases in training loss during the optimization process. However, the validation loss behaviors differ substantially between the random split strategy and the group-based split strategy.

Under the random split strategy, the validation loss remains relatively stable throughout the training process with only minor fluctuations. Although the training loss gradually

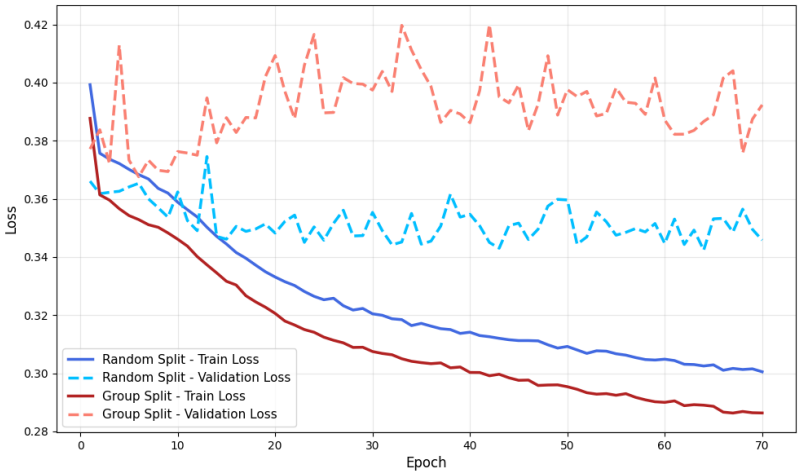


Figure 5.2.: Training and validation loss curves for random sequence splitting and group-based sequence splitting.

decreases, the gap between the training and validation loss remains comparatively small. This behavior suggests that the proposed ConvLSTM model achieved relatively stable learning performance and better generalization capability under the random sequence splitting strategy.

In contrast, the group-based split strategy produced noticeably higher validation losses together with stronger fluctuations during training. Although the training loss continuously decreased and eventually achieved lower values than the random split strategy, the validation loss remained unstable throughout most of the training process. This behavior indicates that the model experienced greater difficulty generalizing to unseen building locations and may have partially overfitted the thermal characteristics of the training dataset.

One possible explanation is that thermal image-based building material classification is strongly influenced by spatial and environmental variations between different buildings. Variations in facade geometry, solar exposure, thermal reflection, weather conditions, and material aging may produce substantial differences in thermal distributions across acquisition locations. Consequently, image sequences collected from different buildings present greater challenges for the proposed ConvLSTM model.

5.3. Seasonal analysis

This study collected thermal images at multiple time intervals between February and April 2026. Variations in ambient temperature, cloud cover, humidity, and solar radiation led to significant differences in the thermal distribution of facades within the dataset.

During the colder period, the thermal contrast of facade surfaces was generally lower, and the spatial temperature distribution was more uniform. In particular, the thermal properties of brick and concrete surfaces were relatively stable due to their high specific heat capacity

and slow heat transfer rate. Under these conditions, the thermal differences between different materials were difficult to distinguish, making material classification from the thermal images challenging.

Conversely, the thermal image sequence acquired during the relatively warmer period showed greater thermal contrast between facade materials. Increased solar radiation and higher ambient temperatures amplified the differences in thermal response between glass, concrete, and brick surfaces. Consequently, the boundaries of some materials became more clearly defined in the thermal images.

5.4. Diurnal analysis

In order to further investigate the influence of daily thermal variation on building material classification, an additional thermal image dataset was collected from the BK building during a single day. The thermal image acquisition was performed at two-hour intervals in order to capture temporal thermal changes under different daytime and nighttime conditions.

The additional dataset consists of 26 image sequences, where each sequence contains six consecutive thermal image frames. Compared with the previous dataset, this acquisition strategy provided denser temporal thermal information and more continuous thermal variation patterns throughout the day.

The evaluation results obtained from the diurnal dataset are summarized in Table 5.2.

Table 5.2.: Evaluation metrics for the diurnal dataset.

Dataset	Class	Precision	Recall	F1-score	IoU
Diurnal	Glass	0.722	0.746	0.734	0.580
	Concrete	0.649	0.662	0.655	0.487
	Brick	0.775	0.712	0.742	0.590

Experimental results obtained from the additional daytime and nighttime dataset demonstrated improved segmentation performance compared with the original thermal image dataset. In particular, higher F1 scores and classification accuracies were achieved for the target building materials.

One possible explanation is that the shorter temporal interval between image acquisitions allowed the proposed ConvLSTM model to capture more consistent thermal evolution patterns of facade materials. Since the thermal variations between consecutive frames became smoother and more temporally correlated, the model was able to learn more stable spatio-temporal thermal features.

Furthermore, the BK building contains relatively diverse facade materials and clear material boundaries, which additionally improved the distinguishability between brick, concrete, and glass regions in thermal imagery.

5.5. Analysis of the proposed hypotheses

Based on the experimental results and thermal image analysis, the proposed hypotheses were generally supported by the observations obtained from the thermal image sequences and the ConvLSTM model.

Different building materials exhibited different thermal behaviors during heating and cooling processes. Brick surfaces showed relatively stable temperature variations and stronger heat retention capability, while glass surfaces demonstrated rapid thermal changes and stronger environmental influence. Concrete surfaces exhibited intermediate thermal behavior between brick and glass.

The experimental results also indicate that different materials produce different spatial thermal distributions. Brick and concrete surfaces generally showed smoother thermal patterns, whereas glass surfaces often exhibited unstable and non-uniform thermal distributions caused by reflections and environmental conditions.

Furthermore, the proposed ConvLSTM model was able to learn meaningful spatio-temporal thermal features from sequential thermal imagery. By utilizing temporal thermal information from multiple frames, the model achieved reasonable segmentation performance for brick, concrete, and glass regions.

However, the results additionally demonstrate that thermal image-based building material classification remains challenging due to environmental variability, thermal reflections, weather conditions, and spatial differences between buildings.

Overall, the proposed ConvLSTM model demonstrated the feasibility of using sequential thermal imagery for pixel-level building material classification and provided useful insights into the thermal behavior differences between facade materials.

6. Discussion, conclusion and future work

6.1. Discussion

Chapter 3 presents the development and optimization of the ConvLSTM model for pixel-level building material classification using spatio-temporal thermal imagery. In Chapter 5, the model was trained and evaluated using different dataset splitting strategies and temporal thermal datasets. The results of this study demonstrate that spatio-temporal thermal sequences combined with the ConvLSTM model are effective for distinguishing facade materials such as glass, concrete, and brick.

The experimental results indicate that temporal thermal behavior plays a significant role in material classification performance. Different building materials exhibit distinguishable heating and cooling patterns due to variations in thermal conductivity, emissivity, heat capacity, and thermal inertia, as discussed in Table 2.1. By incorporating sequential thermal frames, the model can capture both spatial thermal distributions and temporal heat transfer characteristics, providing more discriminative information than single-frame thermal imagery alone.

Based on the random sequence-level split evaluation presented in Table 5.1, the proposed framework achieved a mean f1-score of approximately 0.68 across the three material classes. The findings of this study align with previous research emphasizing the importance of thermal properties in urban and building analysis. Unlike traditional material identification approaches, which mainly rely on RGB imagery or geometric information, the proposed method focuses on dynamic thermal behavior, offering an alternative perspective for facade material characterization. The results suggest that temporal thermal patterns provide valuable complementary information that cannot be fully captured through optical appearance alone.

6.2. Limitation

This section outlines the key limitations in this study, which may influence the robustness, generalizability, and overall performance of the proposed building material classification framework.

Image Registration Sensitivity: The spatio-temporal thermal sequences used in this research require accurate image registration between consecutive frames. However, environmental variations during data acquisition introduced challenges for the registration process. Changes in illumination conditions, reflections, and dynamic background objects such as outdoor blinds, flags, and tree movements affected the consistency between sequential images. These factors may reduce registration accuracy and consequently influence the temporal feature extraction capability of the ConvLSTM model. Since ConvLSTM relies heavily

on temporal consistency across image sequences, registration errors may propagate into the segmentation results and reduce classification performance.

Ground Truth Accuracy: The ground truth masks used for training and evaluation were manually annotated and further refined using the SAM-assisted labeling approach discussed in Section 3.2.2. Since the annotation process relies on manual interpretation of RGB imagery, recognition errors and subjective judgment may occur during material labeling, particularly near material boundaries and mixed-material regions. In addition, inaccuracies in image registration between RGB and thermal imagery may further introduce spatial misalignment errors in the generated thermal labels. These factors may reduce the overall accuracy and reliability of the ground truth dataset.

Limited Temporal Sequence Length and Dataset Size: The dataset used for ConvLSTM training remains relatively limited in both sequence quantity and temporal depth. Each thermal sequence currently contains only five frames, which may not fully capture long-term thermal dynamics and temporal heat transfer characteristics of different building materials. A larger temporal sequence length, for example increasing the number of frames to approximately ten frames per sequence, may allow the model to learn more stable temporal patterns and improve segmentation accuracy. In addition, the overall dataset size is relatively small, which may limit the generalization capability of the model under different environmental conditions.

Environmental and Weather Conditions: The thermal behavior of building materials is strongly influenced by environmental factors such as solar radiation, ambient temperature, humidity, wind speed, and weather conditions. Although this study considered temporal thermal variation, these environmental variables were not explicitly modeled or integrated into the learning framework. Variations in weather conditions may alter thermal signatures and influence classification consistency across different acquisition sessions. Future studies should consider incorporating environmental sensor data and meteorological information to improve model robustness and stability.

In summary, while this study demonstrates the potential of using spatio-temporal thermal imagery combined with ConvLSTM-based semantic segmentation for building material classification, the limitations discussed above highlight several areas for future improvement. Addressing these limitations may improve the reliability, scalability, and generalization capability of the proposed framework for broader urban and building analysis applications.

6.3. Conclusion

This section addresses the research subquestions presented in Section 1.3 and concludes by answering the main research question of this thesis.

- **How can a representative and high-quality training dataset be collected for building material classification using thermal imagery?**

  In this research, the preparation of the training dataset was particularly challenging because pixel-level thermal ground truth data for building materials are not publicly available. To construct the dataset, thermal image sequences were collected using a ground-based thermal camera under different environmental and temporal conditions.

The RGB imagery associated with the thermal sequences was used for manual annotation and SAM-assisted segmentation to generate pixel-level material masks. Subsequently, image registration techniques were applied to align RGB and thermal imagery, allowing the labels to be transferred into the thermal domain. The resulting dataset contains thermal sequences of facade materials including glass, concrete, and brick. Although the generated dataset was sufficient for training and evaluation, limitations such as registration inaccuracies, annotation uncertainty, and restricted material diversity remain challenges for future work.

- **What ConvLSTM model architecture and hyperparameter settings yield the best classification performance for different building materials?**

  Different model configurations and hyperparameter settings were evaluated in Chapter 5. The ConvLSTM model demonstrated stable performance for pixel-level semantic segmentation of building materials. Through experimentation, the optimized configuration included a learning rate of 0.001, a patch size of 64, stride size of 64, and focal loss with class weighting to address class imbalance. In addition, background pixels were excluded during loss computation to reduce the influence of dominant non-material regions. The optimized framework achieved a mean **f1!** (f1!) score of approximately 0.686, indicating that the ConvLSTM-based architecture is capable of learning meaningful spatio-temporal thermal features.

- **How does the temporal variation in thermal imagery (e.g., daily and seasonal changes) impact the classification accuracy of different materials?**

  The experiments demonstrated that temporal variation plays a significant role in thermal material classification. Different materials exhibit distinguishable temporal heating and cooling behaviors due to variations in thermal conductivity, emissivity, thermal inertia, and heat capacity. By incorporating sequential thermal frames, the ConvLSTM model was able to capture both spatial thermal distributions and temporal thermal dynamics. The results indicate that stable environmental conditions and clearer thermal contrasts generally improve classification performance. Furthermore, the study suggests that increasing the temporal sequence length may improve the model's ability to learn more robust thermal patterns and enhance segmentation accuracy.

- **What are the key thermal properties that contribute most to distinguishing between materials in thermal imagery?**

  The results of this study indicate that thermal inertia, heat capacity, emissivity, and spatial thermal distribution are among the most important factors for distinguishing facade materials in thermal imagery. Brick surfaces generally exhibit stable and spatially uniform thermal behavior due to their relatively high thermal inertia, while glass surfaces demonstrate rapid temperature variation and non-uniform thermal distributions caused by reflections and heat transfer characteristics. Concrete exhibits intermediate thermal behavior between brick and glass, making it comparatively more difficult to classify. These thermophysical differences provide valuable discriminative features for spatio-temporal thermal semantic segmentation.

The answers to the research subquestions lead to the conclusion of the main research question:

To what extent is a ConvLSTM-based deep learning model using spatio-temporal thermal imagery suitable for detecting and classifying different building materials and elements?

This thesis demonstrates that a ConvLSTM-based deep learning framework using spatio-temporal thermal imagery is suitable for pixel-level building material classification. The proposed framework successfully captures both spatial thermal distributions and temporal thermal behavior, enabling the differentiation of facade materials such as glass, concrete, and brick.

6.4. Future work

In light of the findings from this research, several directions for future work can be explored to improve the accuracy, robustness, and generalization capability of building material classification using spatio-temporal thermal imagery.

Improvement of Image Registration: One important aspect for future research is the improvement of image registration accuracy between RGB and thermal image sequences. The current registration process is sensitive to environmental variations such as illumination changes, reflections, outdoor blinds, flags, and tree movements. These dynamic background changes may introduce temporal inconsistencies between frames and negatively affect the temporal feature extraction capability of the ConvLSTM model. Future studies could investigate more advanced registration approaches, including deep learning-based registration methods and more robust feature matching algorithms.

Ground Truth Refinement and Annotation Strategies: Future work should also focus on improving the quality and reliability of the ground truth dataset. The current labeling process relies heavily on manual annotation and SAM-assisted segmentation, which may introduce subjective interpretation errors and inaccuracies near material boundaries. More automated and standardized annotation approaches could be explored, such as active learning, interactive segmentation, or multi-modal annotation frameworks combining RGB imagery, thermal imagery, and geometric information. Improving the registration accuracy between RGB and thermal imagery may further enhance the precision of transferred labels.

Expansion of Temporal Sequences and Dataset Size: The temporal sequence length used in this study is relatively limited, with only five frames per thermal sequence. Future research could increase the sequence length to approximately ten frames or more to better capture long-term thermal dynamics and heat transfer behavior of building materials. In addition, expanding the dataset with more acquisition sessions, different buildings, and diverse facade materials would improve the generalization capability of the ConvLSTM framework and reduce the risk of overfitting to specific environments.

Material Diversity and Multi-environment Data Collection: Another promising direction is the collection of thermal imagery from a wider variety of architectural environments and facade materials. The current dataset mainly focuses on glass, concrete, and brick collected from limited acquisition locations. Future studies could incorporate additional materials such as wood, metal panels, plaster, roofing materials, and window frames. Collecting data from different urban environments and architectural styles may provide a more balanced dataset and improve model robustness across unseen scenarios.

Integration of Environmental and Meteorological Data: The thermal behavior of building materials is strongly influenced by environmental factors such as solar radiation, ambient temperature, humidity, wind speed, and weather conditions. Future work could integrate environmental sensor data and meteorological information into the learning framework to

better account for these influences. Incorporating weather-related variables may improve the stability and interpretability of thermal material classification under varying environmental conditions.

Overall, these future improvements may contribute to the development of more robust, scalable, and transferable thermal semantic segmentation frameworks for applications in sustainable building assessment, facade inspection, energy efficiency analysis, digital twins, and urban environmental studies.

A. Declaration of AI/LLM usage

During the preparation of this MSc thesis, Artificial Intelligence (AI) and Large Language Model (LLM) tools were used to support several research and writing-related tasks. The use of these tools is disclosed below in accordance with academic transparency and integrity requirements.

1. **What:**

   The primary AI/LLM tool used in this research was ChatGPT (OpenAI). Additional AI-assisted tools, including GitHub Copilot, were occasionally used during programming and debugging tasks.

2. **How:**

   AI/LLM tools were used to assist with:

   - improving academic writing clarity and grammar;
   - generating and revising LaTeX formatting and table structures;
   - assisting with Python programming, debugging, and code optimization;

3. **Extent:**

   The use of AI/LLM tools was limited to supportive assistance throughout the research and thesis writing process. All scientific analysis, experimental design, implementation, interpretation of results, and final decisions were conducted independently by the author.

4. **Ethics:**

   The author confirms that the use of AI/LLM tools complied with academic integrity guidelines of Delft University of Technology. AI/LLM tools were not used to fabricate data, generate falsified results, or produce plagiarized content. All research findings, experiments, and conclusions presented in this thesis are based on the author's own work and critical evaluation.

B. Reproducibility self-assessment

This appendix provides a self-assessment of the reproducibility of the research presented in this thesis in accordance with open science and research integrity principles.

- **Data Availability:**

  The thermal image dataset used in this research is currently not fully publicly accessible. The dataset consists of manually collected ground-based thermal image sequences of building facades at Delft University of Technology. Due to storage limitations and ongoing dataset organization, the complete dataset has not yet been openly released. However, selected sample thermal image sequences, processed outputs, and example annotations are included in the accompanying project repository.

- **Code Availability:**

  The implementation code, preprocessing scripts, dataset preparation workflow, and model training pipeline developed for this thesis are publicly available through GitHub:

  `https://github.com/Wjy20001/GEO2020_Thesis`

  The repository includes code related to image registration, thermal image preprocessing, patch extraction, ConvLSTM-based semantic segmentation, model training, evaluation, and visualization. The project structure follows reproducibility-oriented development practices recommended by the MSc Geomatics programme. :contentReference[oaicite:0]index=0

- **Documentation:**

  The GitHub repository contains organized scripts, configuration files, and README documentation describing the preprocessing, training, and evaluation workflow. In addition, detailed explanations of the data acquisition process, preprocessing pipeline, model architecture, hyperparameter configuration, and evaluation procedures are provided in Chapter 3 and Chapter 5. These descriptions are intended to support future reproduction and extension of the proposed framework.

- **Reproducibility Rating:**

  The overall reproducibility of this thesis is self-rated as **Moderate**. Most methodological details, model configurations, preprocessing procedures, and evaluation workflows are documented and publicly available through the accompanying repository.

Bibliography

Agarwal C and Sharma A (2011). Image understanding using decision tree based machine learning. doi:10.1109/ICIMU.2011.6122757.

Akiba T, Sano S, Yanase T, Ohta T, and Koyama M (2019). Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD '19, page 2623–2631. Association for Computing Machinery, New York, NY, USA. ISBN 9781450362016. doi:10.1145/3292500.3330701.

Balaras C and Argiriou A (2002). Infrared thermography for building diagnostics. *Energy and Buildings*, 34(2):171–183. ISSN 0378-7788. doi:https://doi.org/10.1016/S0378-7788(01)00105-0. TOBUS - a European method and software for office building refurbishment.

Basha SS, Dubey SR, Pulabaigari V, and Mukherjee S (2020). Impact of fully connected layers on performance of convolutional neural networks for image classification. *Neurocomputing*, 378:112–119. ISSN 0925-2312. doi:https://doi.org/10.1016/j.neucom.2019.10.008.

Baskin C, Liss N, Mendelson A, and Zheltonozhskii E (2017). Streaming architecture for large-scale quantized neural networks on an fpga-based dataflow platform. doi:10.48550/arXiv.1708.00052.

Benedykciuk E, Denkowski M, and Dmitruk K (2021). Material classification in x-ray images based on multi-scale cnn. *Signal, Image and Video Processing*, 15:1–9. doi:10.1007/s11760-021-01859-9.

Chen Y, Chen X, and Chen BM (2026). Enhanced building thermal defect detection using deep learning-based multimodal fusion based thermographic reconstruction. *Journal of Building Engineering*, 120:115473. ISSN 2352-7102. doi:https://doi.org/10.1016/j.jobe.2026.115473.

Choi R, Coyner A, Kalpathy-Cramer J, Chiang M, and {Peter Campbell} J (2020). Introduction to machine learning, neural networks, and deep learning. *Translational Vision Science and Technology*, 9(2). ISSN 2164-2591. doi:10.1167/tvst.9.2.14. Publisher Copyright: © 2020 The Authors.

Desai P, Sujatha C, Chakraborty S, Ansuman S, Bhandari S, and Kardiguddi S (2022). Next frame prediction using convlstm. *Journal of Physics: Conference Series*, 2161(1):012024. doi:10.1088/1742-6596/2161/1/012024.

Dimitrov A and Golparvar-Fard M (2014). Vision-based material recognition for automated monitoring of construction progress and generating building information modeling from unordered site image collections. *Advanced Engineering Informatics*, 28(1):37–49. ISSN 1474-0346. doi:https://doi.org/10.1016/j.aei.2013.11.002.

Fang W, Chen Y, and Xue Q (2021). Survey on research of rnn-based spatio-temporal sequence prediction algorithms. *Journal on Big Data*, 3:97–110. doi:10.32604/jbd.2021.016993.

Gonzalez D, Rueda-Plata D, Acevedo AB, Duque JC, Ramos-Pollán R, Betancourt A, and García S (2020). Automatic detection of building typology using deep learning methods on street level images. *Building and Environment*, 177:106805. ISSN 0360-1323. doi:https://doi.org/10.1016/j.buildenv.2020.106805.

Hu WS, Li HC, Pan L, Li W, Tao R, and Du Q (2020). Spatial–spectral feature extraction via deep convlstm neural networks for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 58(6):4237–4250. doi:10.1109/TGRS.2019.2961947.

Jogin M, Mohana, Madhulika MS, Divya GD, Meghana RK, and Apoorva S (2018). Feature extraction using convolution neural networks (cnn) and deep learning. In *2018 3rd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)*, pages 2319–2323. doi:10.1109/RTEICT42901.2018.9012507.

Junliang C (2022). Cnn or rnn: Review and experimental comparison on image classification. In *2022 IEEE 8th International Conference on Computer and Communications (ICCC)*, pages 1939–1944. doi:10.1109/ICCC56324.2022.10065984.

Kag A and Saligrama V (2021). Training recurrent neural networks via forward propagation through time. In Meila M and Zhang T, editors, *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 5189–5200. PMLR.

Khan A, Sohail A, Zahoora U, and Saeed A (2020). A survey of the recent architectures of deep convolutional neural networks. *Artificial Intelligence Review*, 53. doi:10.1007/s10462-020-09825-6.

Kirillov A, Mintun E, Ravi N, Mao H, Rolland C, Gustafson L, Xiao T, Whitehead S, Berg AC, Lo WY, Dollar P, and Girshick R (2023). Segment anything. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 4015–4026.

Krenker A, Bešter J, and Kos A (2011). Introduction to the artificial neural networks. In Suzuki K, editor, *Artificial Neural Networks - Methodological Advances and Biomedical Applications*, chapter 1. IntechOpen, London. doi:10.5772/15751.

Lai G, Chang WC, Yang Y, and Liu H (2018). Modeling long- and short-term temporal patterns with deep neural networks. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, SIGIR '18, page 95–104. Association for Computing Machinery, New York, NY, USA. ISBN 9781450356572. doi:10.1145/3209978.3210006.

Liu Y, Zhang C, and Dong X (2023). A survey of real-time surface defect inspection methods based on deep learning. *Artificial Intelligence Review*, 56:1–40. doi:10.1007/s10462-023-10475-7.

Ma TY, Li HC, Zheng YB, Du Q, and Plaza A (2025). Fully tensorized lightweight convlstm neural networks for hyperspectral image classification. *IEEE Transactions on Neural Networks and Learning Systems*, 36(8):14213–14227. doi:10.1109/TNNLS.2024.3511575.

Mehmood M, Shahzad A, Zafar B, Shabbir A, and Ali N (2022). Remote sensing image classification: A comprehensive review and applications. *Mathematical Problems in Engineering*, 2022(1):5880959. doi:https://doi.org/10.1155/2022/5880959.

Mehrotra D (2019). *Basics of artificial intelligence & machine learning*. Notion Press.

Mienye I and Jere N (2024). Deep learning for credit card fraud detection: A review of algorithms, challenges, and solutions. *IEEE Access*, PP:1–1. doi:10.1109/ACCESS.2024.3426955.

Mikaél'A B (2013). *Infrared radiation: a handbook for applications*. Springer Science & Business Media.

Ming Y, Cao S, Zhang R, Li Z, Chen Y, Song Y, and Qu H (2017). Understanding hidden memories of recurrent neural networks. In *2017 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pages 13–24. doi:10.1109/VAST.2017.8585721.

Moustapha AI and Selmic RR (2008). Wireless sensor network modeling using modified recurrent neural networks: Application to fault detection. *IEEE Transactions on Instrumentation and Measurement*, 57(5):981–988. doi:10.1109/TIM.2007.913803.

Moyano Campos JJ, Antón García D, Rico Delgado F, and Marín García D (2017). *Threshold Values for Energy Loss in Building Façades Using Infrared Thermography*, pages 427–437. Springer International Publishing, Cham. ISBN 978-3-319-51442-0. doi:10.1007/978-3-319-51442-0_35.

O'Shea K and Nash R (2015). An introduction to convolutional neural networks. *CoRR*, abs/1511.08458.

Patel A, Chatterjee S, and Gorai A (2019). Effect on the performance of a support vector machine based machine vision system with dry and wet ore sample images in classification and grade prediction. *Pattern Recognition and Image Analysis*, 29:309–324. doi:10.1134/S1054661819010097.

Perez H, Tah JHM, and Mosavi A (2019). Deep learning for detecting building defects using convolutional neural networks. *Sensors*, 19(16). ISSN 1424-8220. doi:10.3390/s19163556.

Peters R, Dukai B, Vitalis S, van Liempt J, and Stoter J (2022). Automated 3D reconstruction of LoD2 and LoD1 models for all 10 million buildings of the Netherlands. *Photogrammetric Engineering and Remote Sensing*, 88(3):165–170. doi:10.14358/PERS.21-00032R2.

Pezeshki Z, Soleimani A, Darabi A, and Mazinani S (2018). Thermal transport in: Building materials. *Construction and Building Materials*, 181:238–252. ISSN 0950-0618. doi:https://doi.org/10.1016/j.conbuildmat.2018.05.230.

Pisa I, Morell A, Lopez Vicario J, and Vilanova R (2020). Denoising autoencoders and lstm-based artificial neural networks data processing for its application to internal model control in industrial environments—the wastewater treatment plant control case. *Sensors*, 20:3743. doi:10.3390/s20133743.

Qamar R and Zardari B (2023). Artificial neural networks: An overview. *Mesopotamian Journal of Computer Science*, 2023:130–139. doi:10.58496/MJCSC/2023/015.

Rashidi A, Sigari MH, Maghiar M, and Citrin D (2016). An analogy between various machine-learning techniques for detecting construction materials in digital images. *KSCE Journal of Civil Engineering*, 20(4):1178–1188. ISSN 1226-7988. doi:https://doi.org/10.1007/s12205-015-0726-0.

Rocha J, Santos C, and Póvoas Y (2018). Detection of precipitation infiltration in buildings by infrared thermography: a case study. *Procedia Structural Integrity*, 11:99–106. ISSN 2452-3216. doi:https://doi.org/10.1016/j.prostr.2018.11.014. XIV INTERNATIONAL CONFERENCE ON BUILDING PATHOLOGY AND CONSTRUCTIONS REPAIR, FLORENCE, ITALY, JUNE 20-22, 2018.

Saon G, Tüske Z, Bolanos D, and Kingsbury B (2021). Advancing rnn transducer technology for speech recognition. In *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5654–5658. doi:10.1109/ICASSP39728.2021.9414716.

Savelonas M, Veinidis C, and Bartsokas T (2022). Computer vision and pattern recognition for the analysis of 2d/3d remote sensing data in geoscience: A survey. *Remote Sensing*, 14:6017. doi:10.3390/rs14236017.

Scheepens D, Schicker I, Hlavackova-Schindler K, and Plant C (2023). Adapting a deep convolutional rnn model with imbalanced regression loss for improved spatio-temporal forecasting of extreme wind speed events in the short to medium range. *Geoscientific Model Development*, 16:251–270. doi:10.5194/gmd-16-251-2023.

Schulz H and Behnke S (2012). Deep learning: Layer-wise learning of feature hierarchies. *Künstliche Intelligenz*, 26.

Seo H, Raut AD, Chen C, and Zhang C (2023). Multi-label classification and automatic damage detection of masonry heritage building through cnn analysis of infrared thermal imaging. *Remote Sensing*, 15(10). ISSN 2072-4292. doi:10.3390/rs15102517.

Shanthamallu US and Spanias A (2022). Introduction to machine learning. In *Machine and deep learning algorithms and applications*, pages 1–8. Springer.

Shariq MH and Hughes BR (2020). Revolutionising building inspection techniques to meet large-scale energy demands: A review of the state-of-the-art. *Renewable and Sustainable Energy Reviews*, 130:109979. ISSN 1364-0321. doi:https://doi.org/10.1016/j.rser.2020.109979.

Surendranathan P (2023). *Deep-Learning Enabled Approaches To Raman Spectroscopy - Literature Review*. Ph.D. thesis. doi:10.13140/RG.2.2.34435.78886.

Tuba E, Bačanin N, Strumberger I, and Tuba M (2021). *Convolutional Neural Networks Hyperparameters Tuning*, pages 65–84. Springer International Publishing, Cham. ISBN 978-3-030-72711-6. doi:10.1007/978-3-030-72711-6_4.

Wang L, Xu X, Dong H, Gui R, Yang R, and Pu F (2018). Exploring convolutional lstm for polsar image classification. In *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, pages 8452–8455. doi:10.1109/IGARSS.2018.8518517.

Wang P, Xiao J, Qiang X, Xiao R, Liu Y, Sun C, Hu J, and Liu S (2024). An automatic building façade deterioration detection system using infrared-visible image fusion and deep learning. *Journal of Building Engineering*, 95:110122. ISSN 2352-7102. doi:https://doi.org/10.1016/j.jobe.2024.110122.

Wang Y, Wang N, and Ren Q (2022). Predicting surface heat flux on complex systems via conv-lstm. *Case Studies in Thermal Engineering*, 33:101927. ISSN 2214-157X. doi:https://doi.org/10.1016/j.csite.2022.101927.

Waqas A and and MTA (2024). Deep learning-based segmentation of thermal images for heat loss detection in building façade. *Technology—Architecture + Design*, 8(2):370–379. doi:10.1080/24751448.2024.2405415.

Wellons M (2007). The stefan-boltzmann law. *Physics Department, The College of Wooster, Wooster, Ohio*, 44691:25.

Wilhelm E (2010). *Heat Capacities: Introduction, Concepts and Selected Applications*, pages 1–27. ISBN 978-0-85404-176-3.

Wilson AN, Gupta KA, Koduru BH, Kumar A, Jha A, and Cenkeramaddi LR (2023). Recent advances in thermal imaging and its applications using machine learning: A review. *IEEE Sensors Journal*, 23(4):3395–3407. doi:10.1109/JSEN.2023.3234335.

Yan Q, Jiao F, and Peng W (2024). Building materials classification model based on text data enhancement and semantic feature extraction. *Buildings*, 14(6). ISSN 2075-5309. doi:10.3390/buildings14061859.

Yao G, Lei T, and Zhong J (2018). A review of convolutional-neural-network-based action recognition. *Pattern Recognition Letters*, 118. doi:10.1016/j.patrec.2018.05.018.

Yu Y, Si X, Hu C, and Zhang J (2019). A review of recurrent neural networks: Lstm cells and network architectures. *Neural Computation*, 31(7):1235–1270. ISSN 0899-7667. doi:10.1162/neco_a_01199.

Zafar A, Aamir M, Mohd Nawi N, Arshad A, Riaz S, Alruban A, Dutta A, and Alaybani S (2022). A comparison of pooling methods for convolutional neural networks. *Applied Sciences*, 12:8643. doi:10.3390/app12178643.

Zahiri Z, Laefer DF, and Gowen A (2021). Characterizing building materials using multi-spectral imagery and lidar intensity data. *Journal of Building Engineering*, 44:102603. ISSN 2352-7102. doi:https://doi.org/10.1016/j.jobe.2021.102603.

Zhang G, Pan Y, and Zhang L (2022). Deep learning for detecting building façade elements from images considering prior knowledge. *Automation in Construction*, 133:104016. ISSN 0926-5805. doi:https://doi.org/10.1016/j.autcon.2021.104016.

Zhou S, Rueckert D, and Fichtinger G (2019). *Handbook of Medical Image Computing and Computer Assisted Intervention*. The MICCAI Society book Series. Academic Press. ISBN 9780128165867.

Çekim H, Karakavak H, Özel Kadilar G, and Tekin S (2023). Earthquake magnitude prediction in turkey: a comparative study of deep learning methods, arima and singular spectrum analysis. *Environmental Earth Sciences*, 82. doi:10.1007/s12665-023-11072-1.

Colophon

This document was typeset using LaTeX, using a modified the KOMA-Script class `scrbook` available at https://github.com/tudelft3d/msc_geomatics_thesis_template. The main font is Palatino.