Unsupervised spatio-temporal anomalous thermal behavior monitoring of inside-built environments

Naima Khan[1,2], Md Abdullah Al Hafiz Khan[3], Nirmalya Roy[1,2]
[1]Department of Information Systems, University of Maryland Baltimore County (UMBC),
[2]Center for Real-time Distributed Sensing and Autonomy, UMBC
[3]Department of Computer Science, Kennesaw State University
[1,2](nkhan4, nroy)@umbc.edu, [3]mkhan74@kennesaw.edu

***Abstract*—Continual wavering of outside weather degrades the efficiency of inside building envelope over time and leads to additional energy consumption, various structural damages, etc. Frequent monitoring of the indoor built environment with thermal images can assist in identifying the energy-leaking and potentially damage-prone areas. Although in recent years different researches performed deep learning and computer vision based thermal anomaly detection in built environment, several issues related to conducting strategic non-intrusive indoor thermal inspection using temporal thermal images, are still unresolved in uncontrolled environment of residential buildings. In this work, we propose a scalable thermal image-based monitoring approach for building envelopes combining the visual knowledge of structural joint information among different building components and their corresponding temporal thermal status. We collected longitudinal thermal images from indoor scenes of different building components (e.g., door, window, wall) and employed a high-level spatio-temporal graph (st-graph) to represent the structural connection among different building components and their temporal self-changes. Our proposed novel unsupervised spatio-temporal clustering framework assigns the cluster label to nodes in st-graph, combining its structural (the self and neighboring component) and temporal features which achieves better performance in identifying thermal variation compared to other clustering based approaches. We demonstrate thermal variation across the spots which indicates the potential energy leakage areas inside the built environment. The cluster patterns obtained from our proposed model assist in understanding the thermal characteristics of various surfaces at certain conditions, such as sun reflection and airflow in the inside built environment.**

***Index Terms*—Unsupervised thermal variation, Deep Clustering, Thermal anomalies, Building components, Indoor spaces**

I. Introduction

A sustainable building envelope ensures efficient energy consumption and the thermal comfort of inhabitants inside the built environment. According to a U.S. Energy Information Administration (EIA) study, each household in the USA spends approximately $1500 USD on average for energy utility bills per year [1]. Around 51% of energy consumption in residential buildings contributes to space heating and cooling [2]. Different building components (i.e., walls, windows, doors) contribute to a significant portion of total energy loss due to air leakages, structural deformation, poor insulation materials, etc. From a study on the energy loss through building components, it appears that 35% air leakages are caused by walls, 25% by windows, 25% by roofs or attics, and 15% by floors [3]. Home energy audit expert inspects the inside built environment to identify energy leakage areas and recommend a potential improvement for renovation. Expert-driven leakage area detection requires a thorough frequent inspection of the inside environment, which is time-consuming and expensive. These frequent thermal variation inspection using conventional intrusive equipment is not feasible in residential buildings. Therefore, automatically identifying the potential thermal leakages-prone area helps home energy audit experts quickly inspect the area and recommend necessary improvements and steps to reduce energy usage. Since in and outflow of air through small places can change the surface's thermal condition over time, we consider these sudden thermal behavior changes as thermal anomalies on the built environment surfaces as a proxy for potential energy leakage area detection that could prevent additional energy loss.

Monitoring inside thermal conditions can also help demonstrate the outside environment's impact on analyzing and controlling inside thermal conditions [4]–[6]. Unfortunately, few recent studies focus on thermal variation for a limited number of home spaces to propose theoretical computation of building envelope efficiency and measures the performance of building envelopes in a simulated controlled experimental environment [7]–[9] due to the lack of inbuilt environment longitudinal thermal changes data. Researchers also conduct a field study to explore thermal auditing using smartphones and showcase qualitative analysis of thermal behavioral changes inside the home. This qualitative approach helps analyze thermal changes in conventional thermal leakage-prone areas such as windows and doors, etc.; however, it failed to capture the longitudinal thermal behavior of non-conventional thermal leakage-prone building components (e.g., wall, ceiling). Our work focuses on developing systematic quantitative longitudinal thermal changes (thermal anomaly) of conventional and non-conventional thermal leakage-prone building components in the natural home environment. Developing the quantitative thermal behavioral changes monitoring framework requires systematic data acquisition, scalable system design, integration, and appropriate evaluation of existing IoT-based techniques. Besides, investigating building efficiency requires tuning a large set of parameters (e.g., in and out airflow, inside wind velocity, room pressure, light reflection, etc.) and metadata (e.g., building material types, material decay, etc.), which is often unavailable and inconvenient for detailed inspection. To overcome these challenges, in this work, we perform thermal condition analysis of multiple spaces simultaneously in the absence of building metadata by using the structural information obtained from visual observation.

In the recent past, a limited number of studies explored and quantified building energy leakage from thermal images [10] [11]. [10] detects thermal leakage using threshold techniques, and [11] uses a supervised simplified capsule network to identify thermal leakage using small annotated samples. In this work, we propose a novel unsupervised spatio-temporal graph clustering technique to monitor the thermal variation of different places inside the built environment from smartphone-associated hand-held thermal camera images. We incorporate visual structural connections among building components in identifying thermal variation by presenting the structural relationship among building components with high-level graph representation. Individual objects (e.g., furniture) have their thermal profile that could lead to potential anomaly; hence we ignore the objects around building components in thermal variation analysis, extract spatial and temporal features, and assign cluster labels to each component in the temporal sequence considering their adjacent building component. Our contributions to this work are as follows:

- We identify the thermal anomalies for building components that provide data-driven knowledge about the thermal characteristics of indoor surfaces as a potential indicator of thermal leakage.
- We construct a masked graph from thermal images to capture building components' thermal correlation and propose a novel unsupervised graph-based spatio-temporal deep clustering network-based systematic framework for thermal status monitoring by incorporating structural orientation among different building components from the indoor scene.
- We collected a novel thermal image dataset for an indoor built environment, annotated the dataset with different categories of building components, and evaluated our model performance. To showcase our model's efficacy and effectiveness, we compare our model performance in the presence of state-of-the-art models.

II. Related works

In this section, we mention the relevant works on thermal image based building envelope monitoring and graph based spatio-temporal analysis.

Thermal image based building envelope monitoring has drawn attention to few research groups in recent years. Smart building research area is vastly populated with maintaining thermal comfort, suggesting optimized energy consumption, precise energy disaggregation [12], privacy preservation [13] etc. However, the studies on building components and its energy efficiency monitoring involving IoT based devices, such as, thermal imagers are still not in the mainstream. The key areas of researches include thermal image processing to determine areas of concern, calibration and validation of thermal cameras [14], [15], thermal transmittance using thermal images [16], integration of thermal image monitoring system with HVAC monitoring system [17] as well as case studies in controlled environments [9]. However, in order to identify the damage prone areas in building envelope, both qualitative and quantitative studies are performed using thermal images. Qualitative analysis usually detects the damage prone areas by visual comparison of thermal heatmap in the thermal images while quantitative studies deals the problem from theoretical and computer vision perspectives [11], [18]–[20] in order to detect the commonly known air leakage areas. Unmanned aerial systems equipped with IR cameras is also being used to detect the thermal anomalies using threshold based image processing techniques [10], CAD modeling [21]. However, most of the studies detect thermal anomalies in outside built environment and skipped structural orientation as well as temporal affects of thermal variation inside built environment. In this work, we incorporated structural information to detect thermal anomalies in inside built environment unsupervisedly from longitudinal thermal images of different building components.

Graph based spatio-temporal analysis is performed in many domains to understand the spatially and temporally evolving systems. Several known example includes human motion detection [22], posture detection, reference object tracking in satellite [23] and rgb images [24] and many more. Depending on the domain based problem, several graph based spatio-temporal feature learning architectures are proposed in the literature. Bi et al. proposed graph convolution based architecture for neuromorphic vision sensing which learns composite spatial feature for several tasks such as, classifying objects, event labeling [25]. Spatio-temporal relation between actors and objects are represented by multi-layer dynamic graph representation to detect human activity in video clips [26]. In [27] spatial and temporal graph neural networks are used separately for capturing features from spatial interaction and temporal motion to predict pedestrian trajectories. Jain et al. proposed a spatio-temporal feature learning architecture which provides factorized spatio-temporal graphs for learning features in complex systems [22] like human motion detection. Inspired from this work, in order to demonstrate spatio-temporal thermal variation in building components using their structural connection and temporal thermal evolution, we represent the inside built environment using a high level spatio-temporal graphs and propose a novel unsupervised clustering based thermal anomalies detection framework.

III. Problem formulation

In this section we formulate our thermal variation analysis problem as thermal anomaly detection. We consider thermal anomaly as sudden or abnormal thermal changes on the surfaces of built environment. Usually in and out flow of air through small places can change the thermal condition on the surfaces over time. The goal of our work is to identify the location and time of the thermal anomaly from temporal thermal data of inside built surfaces. Assume I_{B_w} is the set of temporal thermal images $I_0, I_1, \cdots, I_t$ for a building component B_w and we want to extract the subset of images $I_k \in I_{B_w}$ where most thermal variation occurred over the considered period of time. Assume B_d and B_c are two other neighboring building components of B_w. We clustered the set of images I_{B_w} into N_c number of clusters and measures the anomaly score for each image instances. We combine spatial and temporal features of B_w and its neighbors B_d and B_c to assign cluster labels to the instances of I_{B_w}. The cluster assignment function can be expressed as follows:

$$\mathcal{C}_{I^t_{B_{wi}}} \rightarrow f(B_{w_{t_1,\cdots,t_n}}, B_{d_{t_1,\cdots,t_n}}, B_{c_{t_1,\cdots,t_n}}) \quad (1)$$

In the final step, we extract top-k images having higher anomaly score for further qualitative analysis of thermal variation.

IV. Proposed Framework

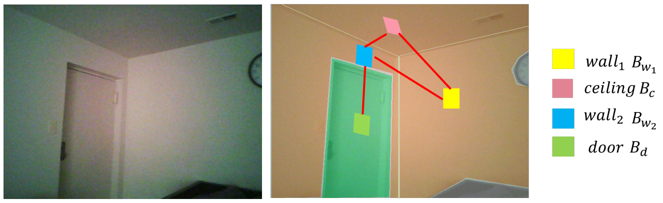
In this section, the principle modules of our proposed framework for thermal status monitoring in inside built monitoring is discussed in detail.

A. Data collection and organization

Previous literature described several issues in collecting and processing longitudinal thermal images, such as privacy concern, prolonged device deployment, maintaining consistency, charge duration etc. However, keeping those issues in consideration, we designed our longitudinal thermal image data collection strategically. Instead of continuous capturing images, we collected images for few minutes in several consecutive hours as surface temperature in inside environment usually changes very slowly. We set the thermal camera on a rotational mount for data collection to cover the desired area in a place for thermal investigation. In this way we can capture thermal images from multiple surfaces altogether simultaneously and relate their thermal variation. In order to deal with the overlapping areas in the captured images, we ask for a human input to select any number of key frames from the captured images so that the frames are completely different to each other. Given a set of key frames and the entire set of images, we perform a simple image retrieval algorithm which finds the top-k most similar images using kNN on the image embeddings with cosine similarity as the distance metric. The image retrieval algorithm is presented in algorithm 1.

Algorithm 1 Image set Construction

```
procedure SIMILAR IMAGE RETRIEVAL(Input: Key Frames $Iq(Iq_1, Iq_2, \cdots, Iq_n)$, $k$, Captured Images $D$ Output: $Sq(Sq_1, Sq_2, \cdots, Sq_n)$)
    $F_q$ ← Features for each of key images $Iq$
    $F_d$ ← Features for each of the images in $D$
    for each $Iq_i$ in $Iq$ do
        $Scores$ ← CosineDistance($Iq_i, D_i$) for $D_i$ in $D$
        Sort the score values in ($Scores$)
        $Sq_i$ ← $D_j$ if $Scores[D_j]$ is in the top k-most images
    return $Sq$
```



(a) Indoor scene

(b) Segmentation of building components

Fig. 1: An example of a indoor scene and corresponding building component annotation

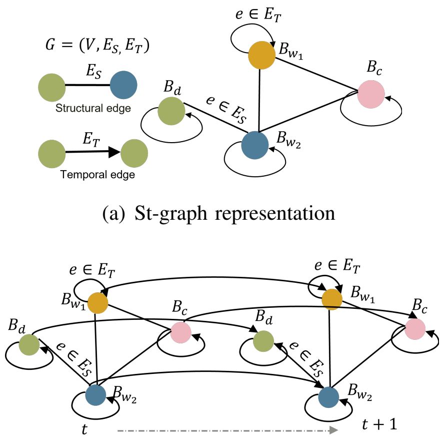
B. Data Pre-processing

In data pre-processing step, we pre-process the retrieved images in the previous step. We extract thermal and visual images of same size from raw MSX (i.e., Multi-spectral dynamic imaging) thermal images. Next, we annotate the segments containing different building components (i.e., wall, ceiling, windows, doors etc.) in the visual images. We also annotate the adjacent joint regions around each building components. Later, we map the temperature values to the image patches consisting of the region of our interest from the corresponding thermal images. We use this temperature map for masking the area in the image. We prepare mask images for each building component present in a single image separately. Then, we compile a temporal series of mask images for each of the building components according to the timestamp associated with original collected images.

C. Spatio-temporal representation of building components

In this subsection, we describe the construction of st-graph with building components and the process of factorized st-graph for our thermal variation analysis in building component .

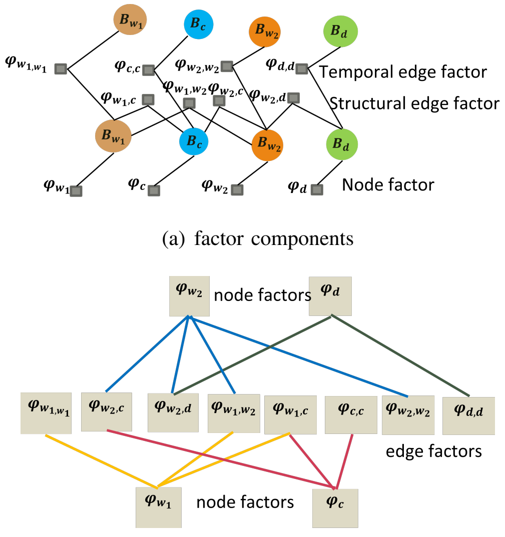
1) St-graph representation: We represent the spatial and temporal connections among each building components using high level spatio-temporal graphs (st-graph). Assume a st-graph $G = (V, E_S, E_T)$ where V represents the building components as nodes, E_S represents structural connection and E_T represents temporal self connection. Figure 1 presents an example of indoor scene with corresponding annotation of different building components. Figure 2(a) shows the st-graph presenting the spatial and temporal connection among different building components (identified in figure 1(b)) for the given indoor scene in figure 1(a). In the compressed st-graph in figure 2(a), temporal edges are showed as loop or self-edges, which refers one building component connected to itself at next timestamp, e.g. B_d at time t connected to B_d at $t+1$. The nodes $v \in V$ and edges $e \in E_S \cup E_T$ repeats over the time for all the images in the dataset. Figure 2(b) shows the same st-graph unrolled over time. The unrolled st-graph shows at a certain time t, nodes are connected to each other with undirected spatial edges $e_s \in E_S$ (e.g. ${B_d}^t$ and $B^t_{w_1}$) while temporal edges $e_t \in E_T$ connects to the node itself at next timestamp (e.g. ${B_d}^t$ and ${B_d}^{(t+1)}$). However, in this



(a) St-graph representation

(b) Unrolled through time

Fig. 2: An example of spatio-temporal graph (st-graph) for a indoor scene



(a) factor components

(b) factor graph

Fig. 3: Factor graph representation of the st-graph (disjoint node factors and edge factors construct a bipartite graph)

work, the purpose of the graph representation is to incorporate the structural connection in the thermal variation analysis of building components. Here, we considered two walls as two different nodes in the st-graph although they are similar kind of building components. We assume the corresponding thermal variation for these two walls would be different from each other as the orientation and the sunlight exposure of these two walls are different.

2) *Simplifying st-graph:* The cluster assignment $\mathcal{C}$ of one node at a certain time depends on the status of this node over a time window as well as the edges associated with it (as expressed in equation 1). In order to represent the complex spatio-temporal system, we use factor graph which simplifies the complex function by introducing multiple simple functions. We present the corresponding factors in figure 3(a) and bipartite factor graph in figure 3(b) for the st-graph showed in figure 2. The factor graph shows the factor functions representing spatial and temporal relations among the building component nodes. As example, the temporal connection between same nodes of B_d are presented by factor function $\phi_{d,d}(\mathcal{C}_d{}^t :\rightarrow B_{d_{t_1,\cdots,t_n}})$ for each node while the structural connection among two different nodes (i.e., B_d and B_{w_1}) presented by pairwise factor function $\phi_{d,w_2}(\mathcal{C}_{dw_2} :\rightarrow B_{dw_2})$ for edge $(B_d, B_{w_2}) \in G$. In the factor graph, one factor is connected to another factor if the corresponding nodes are connected in G. For example, $\phi_{d,d}$ is connected to ϕ_{d,w_2} through ϕ_d.

D. Unsupervised combined feature learning

In this subsection, we propose a deep learning based unsupervised feature learning architecture which combines node and edge features at each time step based on the constructed factor graph and perform more computations on the features for cluster label assignment. Three main modules in the architecture are described as follows.

1) *Encoded node and edge representation:* We used convolution neural network based autoencoder to extract features for each nodes and edges in the st-graph from the corresponding sequential mask images. Assume $\mathcal{I} = \{I_1, I_2, \cdots, I_N\}$ be a set of I mask images for a building component. Mask images contain temperature values for the patches of building components. Autoencoders learns the local features of temperature values by reproducing the similar mask image. Usually autoencoders consist of an encoder and a decoder. In the encoding phase, we reduce the dimension of input data while in the decoding phase, we reconstruct the input data from encoded representation. The reconstruction loss from autoencoder can be represented as follows:

$$L_r = \frac{1}{N_{total}} \sum_{i=1}^{N_{total}} ||I_i - \bar{I}_i||^2 \tag{2}$$

where N_{total} is the total number of images, $\bar{I}_i$ is the reconstructed output of input I_i.

2) *Combine spatial and temporal features:* The factors in the st-graph operate in temporal manner. In order to extract temporal features we use LSTM based recurrent neural networks. We represent each node and edge factor with LSTM based recurrent neural networks to extract the temporal features. We refer the LSTM networks obtained from node factors as nodeLSTMs and the LSTM networks obtained from edge factors as edgeLSTMs. We pass the encoded features of nodes and edges to separate recurrent neural networks in order to capture the temporal latent representations associated with them. The interaction among building components presented in the st-graph are reflected by connections between node LSTMs and the edgeLSTMs. We denote LSTMs corresponding to node factor ϕ_V as $\mathcal{T}_V$ and the edge factor ϕ_E as $\mathcal{T}_E$. In order to obtain a feed forward network, we connect the node and

edge LSTMs following the corresponding node connections in bipartite graph showed in figure 3(b). In particular, the edgeLSTM $\mathcal{T}_E$ is connected to the nodeLSTM $\mathcal{T}_V$ if the factors ϕ_V and ϕ_E are neighbors in st-graph. It refers to the fact that they jointly affect the labels of node in the st-graph. We concatenate node LSTMs and edge LSTMs and the last layer representation from the LSTM network is feed to a clustering layer.

3) Self-supervised clustering: In the clustering module, the LSTMs for node factors learns the temporal representations in self-supervised way presented in [28]. We pre-train the autoencoders and later we train the autoencoders as well as nodeLSTMs and edgeLSTMs end-to-end to assign cluster labels to the corresponding nodes in the st-graph. In clustering module, we compute the similarity between the node representation $\mathcal{B}$ and the cluster vector μ_j using the Student's t-distribution as kernel. Assume q_{ij} is the probability of node B_v to be assigned to cluster j. Hence, the $Q = [q_{ij}]$ is the distribution of the cluster assignments of all samples. After computing the distribution Q of all cluster assignments, we intend to optimize the latent representation of the nodes so that the data representation get closer to the cluster centers and improves cluster coherence. We computed a target distribution P from Q (equation 3) which improves the latent representation generation using the KL divergence loss $\mathcal{L}_c$, between Q and P (equation 4). As P is calculated from Q, using P for updating Q might end up in trivial assignment. Therefore,

$$p_{ij} = \frac{q_{ij}^2/f_j}{\sum_{j' \in K} q_{ij'}^2/f_j'} \tag{3}$$

$$\mathcal{L}_c = KL(P||Q) = \sum_i \sum_j p_{ij} log \frac{p_{ij}}{q_{ij}} \tag{4}$$

we use the distribution of cluster assignments Z from nodeLSTM which can be supervised by target distribution P with following objective function. The KL divergence loss between Z and P, $\mathcal{L}_z$ is computed similarly as in equation 4 and incorporated in the overall loss function. The overall loss function of the proposed architecture is as follows:

$$\mathcal{L} = \mathcal{L}_R + \alpha \mathcal{L}_c + \beta \mathcal{L}_z \tag{5}$$

where $\mathcal{L}_R$ is total reconstruction loss from autoencoders (i.e., nodes and edges), $\alpha, \beta > 0$ are hyper-parameters to balance the clustering optimization preserving the local structure of raw data and to avoid the inference from nodeLSTM in the embedding space. After training for a certain number of epochs when the model reached to a stable condition, we predict the soft assignments in distribution Z.

E. Training

In order to train the architecture, we first the pretrain the autoencoders for each node and edges to obtain the encoded representation. As example, figure 5 shows the forward pass for node B_{w_1}. We feed these encoded representation as features to the LSTM networks for further processing. In forward pass for node B_{w_1}, we have to consider two edges (i.e., $B_{w_1 w_2}$ and $B_{w_1 c}$) associated with this node. The input to edge LSTMs $\mathcal{T}_{w_1 w_2}$ and $\mathcal{T}_{w_1 c}$ is the temporal mask images $\mathcal{I}_{w_1 w_2}$ and $\mathcal{I}_{w_1 c}$, respectively. In the edge LSTMs $\mathcal{T}_{w_1 w_1}$, we pass the sequential encoded representation H_{w_1} for node B_{w_1}. We concatenate layer to layer representations from edge LSTMs $\mathcal{T}_{w_1 w_2}$ and $\mathcal{T}_{w_1 c}$. Later, node LSTM $\mathcal{T}_{w_1}$, at each time step, combines the output of $\mathcal{T}_{w_1 w_1}$ to the concatenated edge features from $\mathcal{T}_{w_1 w_2}$ and $\mathcal{T}_{w_1 c}$. The final output of $\mathcal{T}_{w_1}$ are passed to the clustering module to assign cluster labels to the node. Gradients are updated from end-end to optimize the combined loss from all the modules. The overall training algorithm is presented in algorithm 2 and the comprehensive model architecture is depicted in figure 4.

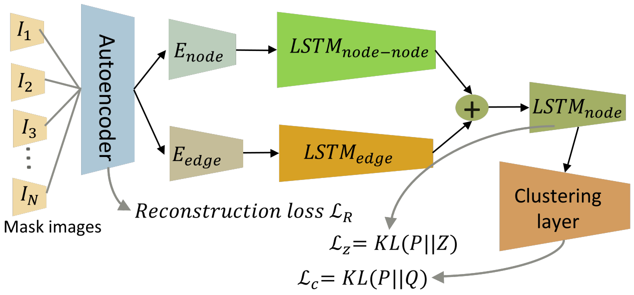


Fig. 4: Model architecture

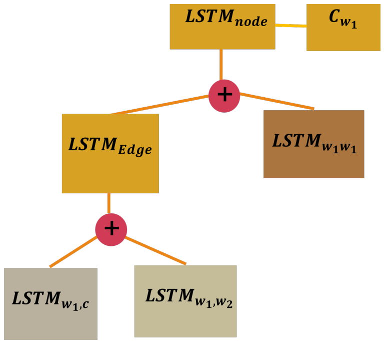


Fig. 5: Feed forward for B_{w_1}

F. Thermal anomaly interpretation

We calculated anomaly scores for each of the images to quantify the deviation of the temperature in the image. For anomaly score computation, we calculated the cosine distance between the image instance and the associated cluster center to it as follows,

$$\text{Anomaly score}(I_i) = \frac{I_i \cdot c_i}{||I_i|| \; ||c_i||} \tag{6}$$

where I_i is the image instance and c_i is the cluster center. We identify the k snaps from each cluster which has higher anomaly scores for each building component to perform further qualitative analysis. In order to justify the anomaly score in interpreting the thermal variation, we use a statistical measure, *Percentage of Anomaly Score Justification (PASJ)*. Assume the number of images show visual thermal variation n_t and total number of image selected by top-k anomaly score

Algorithm 2 Training for unsupervised clustering

```
procedure TRAINING(Input: Graph $G\ (V, E_s, E_T)$, output: $\mathcal{C}_v = c_1, \cdots, c_n$)
    Encode each node: $H_V = H_I, \cdots, H_n$
    Encode each edges: $e \in E_s\ H_e = H_{e_1}, \cdots, H_{e_p}$
    Temporal representation for each node $\mathcal{T}_{V_i V_i} \leftarrow$ edgeLSTM $(H_{V_i})$
    Temporal representation for each edge $\mathcal{T}_{E_i} \leftarrow$ edgeLSTM $(H_{e_i})$
    Concatenate $l$ layer of temporal representation for each pair of edges $\mathcal{T}_{e_i e_j} \leftarrow H^l_{e_i} \oplus \mathrm{H}^l_{e_j}$ where $(e_i, e_j) \in E_s$
    Connect $\mathcal{T}_{V_i} \oplus \mathcal{T}_{e_i e_j}$ in $T_{VE}$
    Represent $\mathcal{T}_{VE}$ with nodeLSTM $\mathcal{T}_V$
    Compute $z$ with ClusteringLayer($\mathcal{T}_{VE}$)
    Compute cluster labels $\mathcal{C}$ from $z$
    Return $\mathcal{C}_v = c_1, \cdots, c_n$
```

N_k, the percentage of top-k anomaly score justification P_k is as follows:

$$P_k = \frac{n_t}{N_k} * 100\% \tag{7}$$

V. Experimental Setup

In this section, we present our data acquisition, pre-processing techniques as well as quantitative evaluation and qualitative analysis of the thermal status of various building components applying our proposed framework with two case studies.

A. Data acquisition

We collected thermal images of indoor built environment using a android smartphone associated FLIROne thermal camera which provides both the RGB and thermal images. Thermal resolution of the collected images are 640×480. We deployed the hand-held smartphone associated thermal camera on a rotating mount placed on top of a height adjustable tripod as depicted in figure 6. Therefore, we can cover the maximum area we want to investigate as well as capture images of multiple surfaces simultaneously. For our data collection, we captured the images from 1-1.5 meter distance from all the considered surfaces. We also developed an android application to collect longitudinal thermal images which captures images in every 10 seconds. We collected data in consecutive 4-5 hours at different time of several days from two indoor scenes. Each hour we captured images for 10 minutes by 180° camera rotation. Figure 7(a) provides the visual frames of one rotation of the camera. We collected approximately 5000 thermal images in about 15 hours of data collection for each of indoor scenes.

B. Data pre-processing

We pre-process the raw MSX thermal images by extracting thermal and visual images. We extract temperature values in celsius unit. However, extracted thermal images are one-channel temperature values associated with the surface. Later, we identify the building components of our interest and select the key frames from the pool of collected images where key frames have no overlapping areas among them. This helps in identifying the set of images similar to key frames which consist of building components of our interest. Figure 7(a) shows the captured images in one rotation of the camera and the selected key frames by human selection are highlighted with red box in figure 7(b).

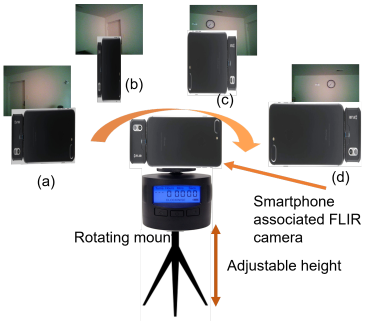
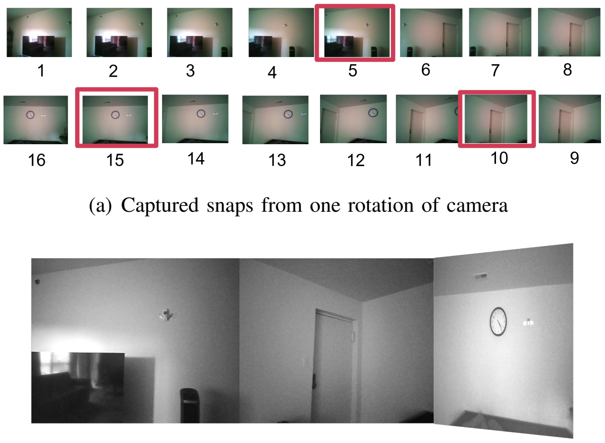


Fig. 6: Overall setup for data collection. While keeping the smartphone thermal camera on a rotating mounting (360 degrees) to capture thermal images of different example rotational positions (a), (b), (c), (d).

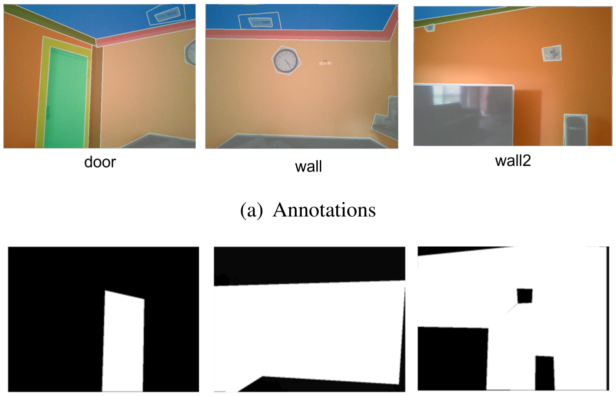


(a) Captured snaps from one rotation of camera

(b) View of the scene by combining key frames

Fig. 7: Example of captured images and summary scene

In next phase, we annotated the image patches which contains building components. Figure 8(a) shows an example annotation for the three images of indoor scene-I. Another scene considered in our experiment is showed in figure 10. However, in order to prepare the temperature mask images of building components, we exclude the pixels of other objects and components inside of the concerned component. In that way, we prepare the list of temporal mask images for each building components and arrange them in a sequence according to the associated timestamp. As example, the middle image in figure 7(b) shares the walls both from left and right side. The list of mask images for one of the walls contains the corresponding wall patches from all the snaps and align them according to the snap timestamp. We map these image patches



(a) Annotations

(b) Mask images of building components

Fig. 8: Annotated area of interests and corresponding mask images

to temperature values from the corresponding thermal image. The annotated building components and corresponding masks for building components (i.e., door and two walls) in the scene frames (presented in figure 7(b)), are showed in the figure 8(a) and (b), respectively.

C. Implementation details

The experiments are conducted on a Linux server integrated with Intel i7-6850K CPU, 4x NVIDIA GeForce GTX 1080Ti GPUs and 64GB RAM. All the codes of data preprocessing, visualization and deep learning algorithms are implemented with Python. Especially for deep learning, PyTorch libraries are used. In order to pre-process the large amount of image data parallel processing, and trained the entire network by distributing the data among two gpus.

We used three blocks of convolution and maxpooling layer for the encoding layers of autoencoder. The kernel filter size for three convolution layer is (4,4), (4,6) and (2,2). After the third convolution block in the encoder, the output is flattened and used as the encoded representation of the mask images. In the decoder, the filter sizes are reversed. Encoded output of node and edges are passes through three LSTM layers. The dimension for the node LSTMs are obtained by dividing the output size of three encoding layers for the node by 2. The dimension for the edgeLSTMs are derived from the encoded output size of the edges. We obtained the first, second and third lstm layer output dimension by dividing the encoded output dimension of edge by 4, 8 and 16 respectively. We pretrained the autoencoders for 100 epochs and trained the clustering module for 60 epochs on average until the cluster assignments stops changing.

D. Evaluation

We evaluate the proposed model with two other relevant base models i.e., DEC [29] and LSTM-3lr [30] of deep clustering using our collected dataset. Here, we select these two base deep clustering models as one of them i.e., DEC uses spatial features and the other one i.e., LSTM 3-lr uses temporal features. This helps us in understanding the advantage of using both spatial and temporal features together for deep clustering as well as incorporating clustering loss to construct the encoded representation. The brief description of two base models are as follows:

Deep Embedded Clustering (DEC) introduces the concept of simultaneous optimization of both data representation at lower dimensional space and clustering loss [29].

LSTM 3-lr proposed by Ghosh et al. combines a three layer LSTM network with a dropout autoencoder to capture temporal features of human posture [30] to predict human motion over a long period of time . We deployed this architecture by using a drop-out autoencoder for extracting spatial features for mask images and later filter the prediction by each building component the same architecture.

We use two clustering metrics i.e., NMI, ARI to evaluate our proposed unsupervised spatio-temporal clustering for thermal anomaly detection. Besides, from qualitative analysis we also provide empirical measure, PASJ which provides the percentage of top-k extracted images having visual appearance of thermal variation. This score provides the idea of statistically how much we can rely on the finally extracted snaps for observation.

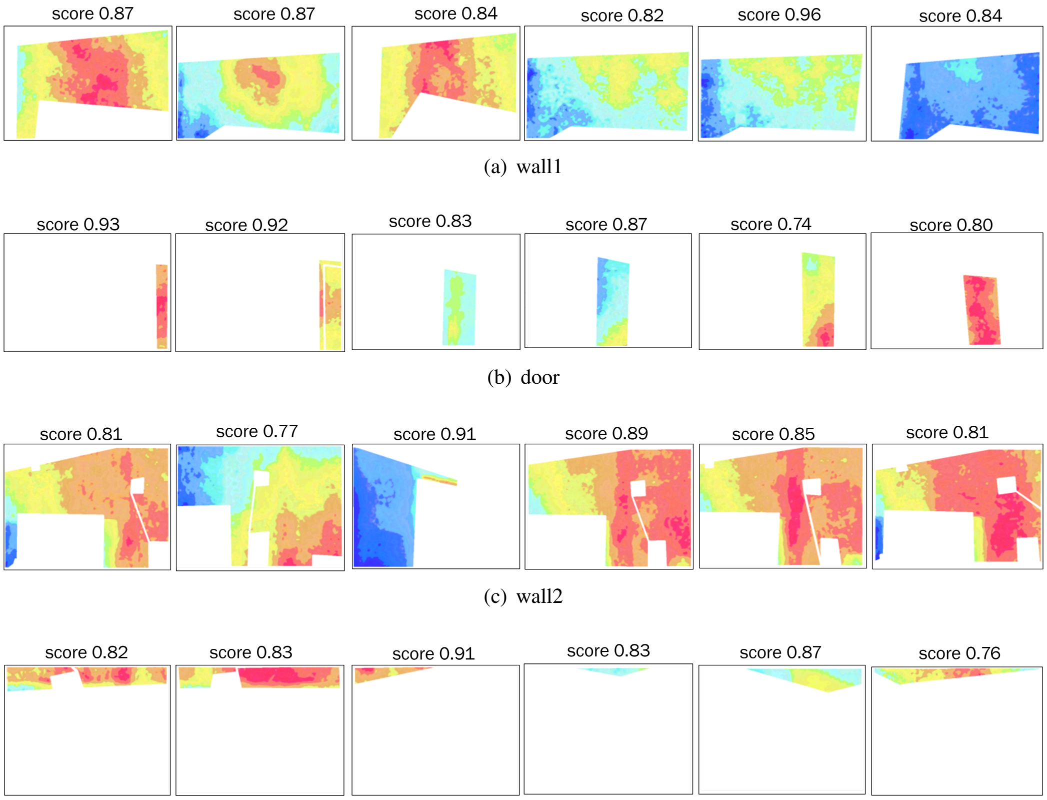
We present the evaluation results from two baseline models and the proposed model in table I. Our proposed model achieves better cluster with temporal building components than the other two models. From the qualitative analysis, we observe that DEC assigns different cluster label to visibly similar thermal variation in the sequence while LSTM 3-LR deals same building components from overlapping snaps similarly and performs poorly for the components having smaller area, like, ceiling and ventilators in our scenarios.

TABLE I: Evaluation metrics

Metrics	DEC		LSTM-3LR		Proposed	
	scene-I	scene-II	scene-I	scene-II	scene-I	scene-II
NMI	0.467	0.452	0.569	0.558	0.799	0.781
ARI	0.697	0.651	0.478	0.450	0.745	0.727
PASJ	69.7%	65.1%	47.8%	45.0%	70%	72%

E. Thermal variation analysis

Here we present the qualitative analysis of thermal variation for indoor scenario presented in figure 7 and 10. We present the thermal status of four building components, i.e., door, two walls on the left and right of the door and ceiling in the scene. Figure 9 shows the thermal gradient for these building components along with the associated anomaly score calculated from three clusters separately. We visualized the thermal status for each of the building components having highest two anomaly scores from each three clusters. The first two columns (i.e., i and ii) in the figure show the images from cluster $\mathcal{C}_1$, next two columns (i.e., iii and iv) show images from cluster $\mathcal{C}_2$, and the last two columns (i.e., iv and v) for $\mathcal{C}_3$.



(a) wall1

(b) door

(c) wall2

(d) ceiling

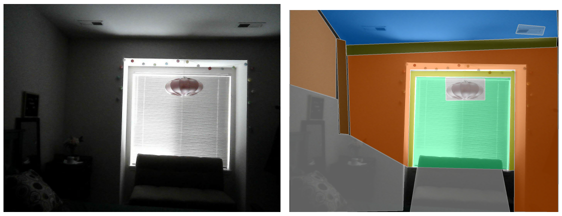
Fig. 9: Thermal variation over building components for indoor scene-I

Scene-I: We can observe the patterns of thermal changes over the places in the first indoor scene from figure 9(a)-(d) as follows.

Walls: In figure 9(a) shows the sparse thermal change on the left side wall (i.e., $wall_1$) of the door. The images in the columns i-iv of this row show higher intensity in the middle which occurs due to sun reflection on $wall_1$ from the opposite window. In the last two column v-vi, higher intensity of cooler temperature can be noticed for the wind flow through the door and window from two sides of this wall. Figure 9(c) shows the thermal status change for $wall_2$. Similar to $wall_1$, sun reflection from the opposite window and wind flow through the door explains higher intensity of warmer and cooler tone on the right and left side of the wall, respectively.

Door: In figure 9(b)(i-vi), we notice the air flow from all four sides of the door through door gaps. We can observe higher intensity of thermal change through top and bottom gap of the door.

Ceiling: Thermal changes for ceiling portion from three images (figure 7) are showed in figure 9(d). The ceiling portion in the first four columns i-iv comes from the leftmost image of the scene 9 where we observe higher intensity of thermal change in the middle portion of ceiling. The images in the last two columns v-vi of this row, comes from the rightmost image of the scene in figure 7. Thermal bridge or the sunlight reflection might cause the higher intensity in this area.

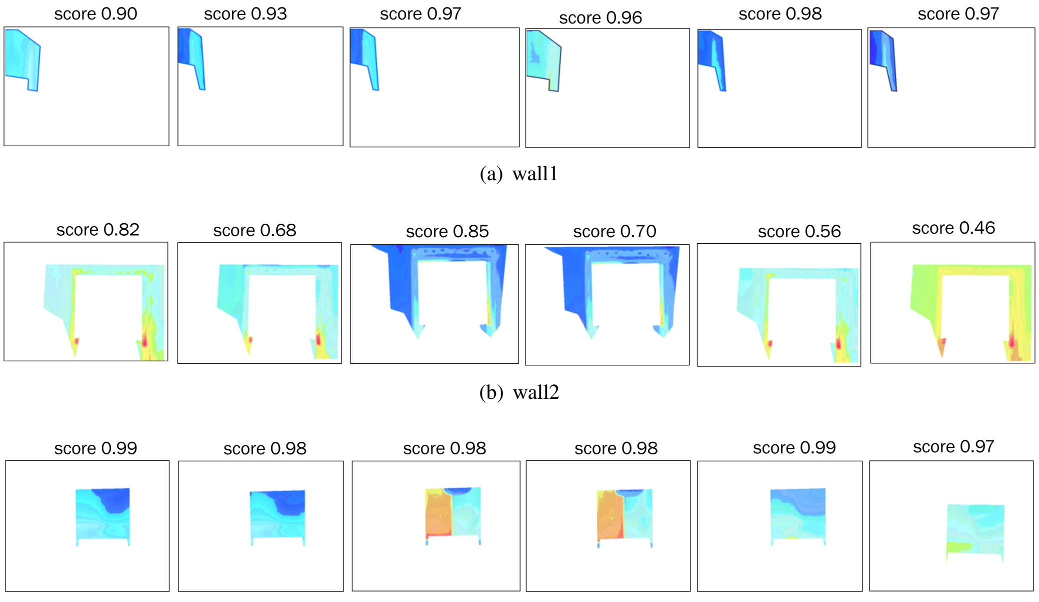


(a) visual of indoor scene

(b) Annotations

Fig. 10: Indoor scene-II and annotation of corresponding building components

Scene-II: In the second scenario, we deploy our framework for another indoor scene presented in figure 10(a). In this case, we placed the camera in a static position and collected the thermal images for spot longitudinally. The annotation of the building components for this scene are depicted in figure 10(b). We consider four building components (i.e., window, left and right side wall of the window) for the scene. The dataset was collected in a snowy day with open and closed window condition. Therefore, the thermal status of different building components in this figure shows cooler tone. However, we can visually observe the impact of extreme cold weather on the inside surfaces around the window area, presented in figure 11(a)-(c). Similarly, the snaps with top-2 highest anomaly



(a) wall1

(b) wall2

(c) window

Fig. 11: Thermal variation over building components for indoor scene-II

scores from three clusters are presented here. The thermal changes over a time period is showed in the figure. We can notice different intensity of cooler tone on the surface.

Walls: In figure 11(a) and (b) we can observe thermal changes over the adjacent walls of the window. For the left side wall (i.e., $wall_1$), figure 11(a) shows very similar thermal changes for each of the clusters. For the wall around the window area, the images in the columns i-iv of figure 11(b) show higher intensity. In the last two column v-vi, we observe similar temperature over the entire area except few spots.

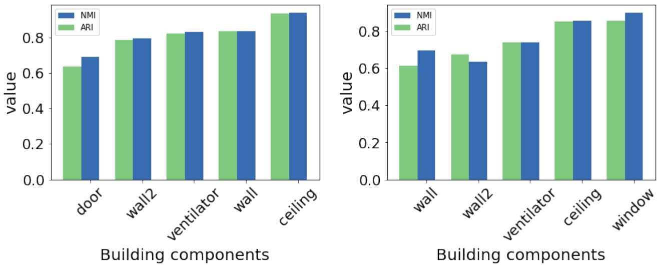
Window: Figure 9(c)(i-ii) and (v-vi) show the thermal status of the window when the left part of it kept open. We observe higher intensity of cooler tone in the upper right corner of the window in i-ii of this figure. For figure 9(c)(iii-iv) represents the window status when its left side is covered.

F. Individual building component

We present the NMI and ARI for each building component in two indoor scenes in figure 12. We observe the clustering metrics for wall components are relatively lower than the other components, as the temperature on wall surfaces are more sparse than that of others. The clusters for smaller area like ceiling and ventilator achieves better cluster metrics which indicates the differentiation of subtle thermal changes over these area.

VI. Discussion

Our proposed thermal anomaly detection framework, provides the location and time of the thermal anomalies for different building components in inside built surfaces from unsupervised clustering approach. From spatio-temporal perspective, as it is difficult to define the ground truth of thermal anomalies, we approach this problem from unsupervised



(a) Indoor scene I

(b) Indoor scene II

Fig. 12: Cluster metrics obtained from target distribution

manner. We considered detecting the sudden thermal changes in surfaces like wall along with the conventional spaces i.e., around the door and window. We connect the thermal status of adjacent door and windows through high level spatio-temporal graphs to determine the thermal status of a building component like walls and ceiling. However, the construction of st-graph may vary upon how much contextual factors we want to consider. In our experiments, we only considered the visible structural connection among building components. As we followed unsupervised approach in our work, we could not evaluate the approach with other metrics, such as, AMI and mIoU used in image segmentation based thermal anomaly detection approach [11]. In this work, we skipped tracking air flow in indoor scenarios which would be beneficial to understand the thermal comfort of inhabitants in more practical way. In the future extension of this work, we plan to extend our experiment to the indoor scenarios with more building components in order to evaluate the robustness of our proposed model.

VII. Conclusion

In this work, we proposed a end-to-end framework for systematic thermal evaluation of inside built environment using low-cost non-intrusive IoT-devices without metadata of buildings. Our graph representation of spatio-temporal thermal correlation among building components assist in understanding the non-trivial thermal variation on building indoor surfaces as well as identifying the location of energy leakages which can reduce energy consumption and prevents severe damages beforehand.

Acknowledgment

This work has been supported by U.S. Army grant W911NF2120076 and ONR grant N00014-18-1-2462.

References

[1] "Energy bills — department of energy information administration," 2021.

[2] "Use of energy in homes — department of energy information administration," 2015.

[3] "Energy loss in homes and the benefits of insulation [infographic]: How to insulate a home buying guide to home insulation," 2013.

[4] M. K. Singh, S. Mahapatra, S. Atreya, and B. Givoni, "Thermal monitoring and indoor temperature modeling in vernacular buildings of north-east india," *Energy and Buildings*, vol. 42, no. 10, pp. 1610–1618, 2010.

[5] N. Khan, M. Ahmed, and N. Roy, "Temporal clustering based thermal condition monitoring in building," *Sustainable Computing: Informatics and Systems*, vol. 29, p. 100441, 2021.

[6] N. Khan and N. Roy, "Builtnet: Graph based spatio-temporal indoor thermal variation detection," in *2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pp. 1696–1703, IEEE, 2021.

[7] W. Liu, X. Zhao, and Q. Chen, "A novel method for measuring air infiltration rate in buildings," *Energy and Buildings*, vol. 168, pp. 309–318, 2018.

[8] J. L. Lerma, M. Cabrelles, and C. Portalés, "Multitemporal thermal analysis to detect moisture on a building façade," *Construction and Building Materials*, vol. 25, no. 5, pp. 2190–2197, 2011.

[9] E. Barreira, R. M. Almeida, and M. Moreira, "An infrared thermography passive approach to assess the effect of leakage points in buildings," *Energy and Buildings*, vol. 140, pp. 224–235, 2017.

[10] B. Kakillioglu, S. Velipasalar, and T. Rakha, "Autonomous heat leakage detection from unmanned aerial vehicle-mounted thermal cameras," in *Proceedings of the 12th International Conference on Distributed Smart Cameras*, pp. 1–6, 2018.

[11] C. Pan, J. Wang, W. Chai, B. Kakillioglu, Y. El Masri, E. Panagoulia, N. Bayomi, K. Chen, J. E. Fernandez, T. Rakha, *et al.*, "Capsule network-based semantic segmentation model for thermal anomaly identification on building envelopes," *Advanced Engineering Informatics*, vol. 54, p. 101767, 2022.

[12] B.-J. Ho, H.-L. C. Kao, N.-C. Chen, C.-W. You, H.-H. Chu, and M.-S. Chen, "Heatprobe: a thermal-based power meter for accounting disaggregated electricity usage," in *Proceedings of the 13th international conference on Ubiquitous computing*, pp. 55–64, ACM, 2011.

[13] P. R. Ovi, E. Dey, N. Roy, A. Gangopadhyay, and R. F. Erbacher, "Towards developing a data security aware federated training framework in multi-modal contested environments," in *Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications IV*, vol. 12113, pp. 189–198, SPIE, 2022.

[14] M. L. Mauriello, M. Saha, E. B. Brown, and J. E. Froehlich, "Exploring novice approaches to smartphone-based thermographic energy auditing: A field study," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pp. 1768–1780, 2017.

[15] M. L. Mauriello, J. Chazan, J. Gilkeson, and J. E. Froehlich, "A temporal thermography system for supporting longitudinal building energy audits," pp. 145–148, ACM, 2017.

[16] R. Albatici, A. M. Tonelli, and M. Chiogna, "A comprehensive experimental approach for the validation of quantitative infrared thermography in the evaluation of building thermal transmittance," *Applied energy*, vol. 141, pp. 218–228, 2015.

[17] V. Tanasiev, G. C. Pătru, D. Rosner, G. Sava, H. Necula, and A. Badea, "Enhancing environmental and energy monitoring of residential buildings through iot," *Automation in Construction*, vol. 126, p. 103662, 2021.

[18] G. Park, M. Lee, H. Jang, and C. Kim, "Thermal anomaly detection in walls via cnn-based segmentation," *Automation in Construction*, vol. 125, p. 103627, 2021.

[19] J. Yang, W. Wang, G. Lin, Q. Li, Y. Sun, and Y. Sun, "Infrared thermal imaging-based crack detection using deep learning," *Ieee Access*, vol. 7, pp. 182060–182077, 2019.

[20] H. Perez, J. H. Tah, and A. Mosavi, "Deep learning for detecting building defects using convolutional neural networks," *Sensors*, vol. 19, no. 16, p. 3556, 2019.

[21] T. Rakha and A. Gorodetsky, "Review of unmanned aerial system (uas) applications in the built environment: Towards automated building inspection procedures using drones," *Automation in Construction*, vol. 93, pp. 252–264, 2018.

[22] A. Jain, A. R. Zamir, S. Savarese, and A. Saxena, "Structural-rnn: Deep learning on spatio-temporal graphs," in *Proceedings of the ieee conference on computer vision and pattern recognition*, pp. 5308–5317, 2016.

[23] F. Guttler, D. Ienco, J. Nin, M. Teisseire, and P. Poncelet, "A graph-based approach to detect spatiotemporal dynamics in satellite image time series," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 130, pp. 92–107, 2017.

[24] L. Khiali, M. Ndiath, S. Alleaume, D. Ienco, K. Ose, and M. Teisseire, "Detection of spatio-temporal evolutions on multi-annual satellite image time series: A clustering based approach," *International Journal of Applied Earth Observation and Geoinformation*, vol. 74, pp. 103–119, 2019.

[25] Y. Bi, A. Chadha, A. Abbas, E. Bourtsoulatze, and Y. Andreopoulos, "Graph-based spatio-temporal feature learning for neuromorphic vision sensing," *IEEE Transactions on Image Processing*, vol. 29, pp. 9084–9098, 2020.

[26] M. Tomei, L. Baraldi, S. Calderara, S. Bronzin, and R. Cucchiara, "Video action detection by learning graph-based spatio-temporal interactions," *Computer Vision and Image Understanding*, vol. 206, p. 103187, 2021.

[27] H. Zhou, D. Ren, H. Xia, M. Fan, X. Yang, and H. Huang, "Ast-gnn: An attention-based spatio-temporal graph neural network for interaction-aware pedestrian trajectory prediction," *Neurocomputing*, vol. 445, pp. 298–308, 2021.

[28] D. Bo, X. Wang, C. Shi, M. Zhu, E. Lu, and P. Cui, "Structural deep clustering network," in *Proceedings of The Web Conference 2020*, pp. 1400–1410, 2020.

[29] J. Xie, R. Girshick, and A. Farhadi, "Unsupervised deep embedding for clustering analysis," in *International conference on machine learning*, pp. 478–487, PMLR, 2016.

[30] P. Ghosh, J. Song, E. Aksan, and O. Hilliges, "Learning human motion models for long-term predictions," in *2017 International Conference on 3D Vision (3DV)*, pp. 458–466, IEEE, 2017.