End-to-end Model for Estimating Heat loss of Building envelope using Deep learning with Infrared image

Heesung Park1†, Sangho Son1†, Ji Young Kang1†, Haemin Jung1, Seungeon Lee2, Deuk-Woo Kim2 and Wooju Kim1*

1 Department of Industrial Engineering, Yonsei University, Seoul, Republic of Korea
2 Korea Institute of Civil Engineering and Building Technology, Goyang-Si, Republic of Korea

Abstract
Estimating heat loss on the envelope of buildings is important in efficient management of energy within buildings. To determine the degree of heat loss, various methods which use infrared images as input data have been suggested. However, the studies have limitations; they require consideration of environmental conditions and lots of additional metadata inputs to describe them. This paper presents a novel method structured as an end-to-end process using a convolutional neural network (CNN) to estimate heat loss, only with infrared images of the building envelope. The process is divided into three steps: CNN-based object detection to identify building components from infrared images, edge detection to calculate actual area, and estimation of relative heat loss among objects with identical material. The results of our process are estimated heat loss types of the objects with their classes and locations. We first generate a temperature distribution using a histogram of pixel colors of object areas, decide a threshold to determine whether the image shows heat loss or not and assign heat loss types based on the threshold. We experimented our method for windows and showed its usability. In the future, we will generalize our method by applying it to other parts of building envelope.

Key Innovations
- We proposed an end-to-end model capable of object detection to heat loss estimation.
- We proposed a CNN-based object detection model for infrared images, trained from real-world images relatively easy to be collected.
- We suggested a methodology of heat loss estimation by generating temperature distribution. It can estimate heat loss with less metadata.

Introduction
Problem Definition
Excessive generation of carbon dioxide (CO2) which derives recent climate change has been a major issue for decades, and buildings are taking a big part of this problem. The carbon dioxide emitted from buildings accounts for about 40% of global amount, and it comes out in the process of generating a large amount of thermal energy, about 36% of the total energy consumption. (IEA 2018) For the purpose of reduce thermal energy production and eventually the amount of emitting carbon dioxide, efficient thermal energy management of buildings is essential. In addition, for the efficiency of such building management, estimating heat loss of building envelope is important to eliminate or minimize heat loss through the building envelope (O’Grady et al., 2017, Najjar et al., 2019).

In order to estimate the thermal energy efficiency of the building, the thermal insulation performance has been estimated. As a quantitative measure, thermal transmittance value is generally used (Nardi et al., 2014). Albaciti et al. (2010) proposed a representative method using the thermal transmittance value which uses surface temperature obtained by an infrared camera and environmental variables such as emissivity and wind speed. But there is a difficulty in having a suitable environment; photographing and measurement are required under certain conditions (Choi et al., 2017). In other words, the method for estimating the thermal energy efficiency using thermal transmittance has limitations due to limited situation, difficulties requiring various input values, and difficulties in real-time judgment.

Proposed Method Overview
This study proposes a process that has the following three strengths.
1. It is easy to measure and judge heat loss by using only infrared images as inputs.
2. It is less affected by differences in environmental conditions in the image in that there are the same environmental conditions and that the components proceed with the same domain-specific comparisons.
3. It provides an objective estimation that analyzes the heat loss by quantifying the temperature distribution on the detected object.

† Heesung Park, Sangho Son, and Ji Young Kang contributed equally to this work.
* Corresponding author
Calculating Object Surface Area based on Detected Edges

In the second step, edge detection is conducted on identified objects and its surface area is calculated. Common edge detection methods include Canny edge detection (Canny, 1986) and holistically-nested edge detection (Xie et al., 2015). The Canny method uses Gaussian filters to remove noise from images and then detects objects’ edges by calculating the rate of change in pixel color. The holistically-nested edge detection model utilizes a neural network and identifies objects’ edges mainly based on their outer contours. However, it does not work well for small amounts of data and does not work quickly and so was not suitable for the proposed method.

We included the Canny edge detection method in the proposed process. The CNN then identifies the point on the edges with HoughLinesP (Matas et al, 2000). Then it identifies the closest point to the edge of the image to determine the identified object’s surface area.

Reason of choosing Object and Edge Detection

Among the tasks of object recognition, there are object detection, which finds the position of an object in a box shape (Figure 2(a)), and segmentation, which finds an object in units of pixels (Figure 2(b)). So, one may question whether it is possible to replace the application of edge detection after object detection with a segmentation model.

Object Detection in Infrared images

In the first step, a deep learning-based object detection model is used. Based on labeled image dataset, a neural network is trained to produce an output similar to a ground truth. And when a new image comes in, it finds an object by using the trained network. This object detection finds objects in an image or video, identifies their type, and detects the positions in the form of a rectangular box. This is a key technology used in all the industries such as auto-driving, facial recognition, video surveillance, and Internet of Things (IoT).

Representative object detection models are Faster R-CNN (Ren et al., 2015), SSD (Liu et al, 2016), and YOLO (Redmon et al., 2016). Faster R-CNN first identifies an area where an object might be located and then identifies the object. It performs the two deep neural network operations during training, so its training and processing speeds are relatively slow. SSD is a single network and is faster and more accurate than Faster R-CNN, but it is difficult to use.

The YOLO-V4 (Bochkovskiy et al., 2020), the SOTA version of YOLO, was the CNN used in the proposed process because it identifies objects in infrared images quickly enough for real-time use and is easy enough to train that it can be trained on a single graphics processing unit.

The following sections describe each step in more detail, why certain choices were made in designing the proposed model, and related works are described.

- **Object Detection in Infrared images**: In infrared images, it detects the types and locations of objects, such as windows and doors, that have relatively high heat loss from buildings.

- **Calculating Object Surface Area based on Detected Edges**: Use edge detection within the location where the object was detected to estimate the exact proportion and size of the actual object.

- **Heat Loss Estimation Mechanism**: The temperature distribution within the detected object is quantified and relative analysis is performed. Finally, outputs are the class, location of objects, and the amount of heat loss as type.

The objects are then segmented to identify their shape, and their surface areas are calculated. This is done by summing the areas of the detected objects and subtracting the overlapping areas. This process is repeated for each object detected in the image.

The following sections describe each step in more detail, why certain choices were made in designing the proposed model, and related works are described.

Figure 1) Proposed process diagram.

As a result, it provides an end-to-end process (i.e., a fully functional system without additional intervention) as shown in Figure 1 to facilitate heat loss estimation. The proposed process occurred in three steps.

- **Object Detection**
- **Calculating Object Surface Area**
- **Heat Loss Estimation Mechanism**

The following sections describe each step in more detail, why certain choices were made in designing the proposed model, and related works are described.

**Object Detection in Infrared images**

In the first step, a deep learning-based object detection model is used. Based on labeled image dataset, a neural network is trained to produce an output similar to a ground truth. And when a new image comes in, it finds an object by using the trained network. This object detection finds objects in an image or video, identifies their type, and detects the positions in the form of a rectangular box. This is a key technology used in all the industries such as auto-driving, facial recognition, video surveillance, and Internet of Things (IoT).

Representative object detection models are Faster R-CNN (Ren et al., 2015), SSD (Liu et al, 2016), and YOLO (Redmon et al., 2016). Faster R-CNN first identifies an area where an object might be located and then identifies the object. It performs the two deep neural network operations during training, so its training and processing speeds are relatively slow. SSD is a single network and is faster and more accurate than Faster R-CNN, but it is difficult to use.

The YOLO-V4 (Bochkovskiy et al., 2020), the SOTA version of YOLO, was the CNN used in the proposed process because it identifies objects in infrared images quickly enough for real-time use and is easy enough to train that it can be trained on a single graphics processing unit.
Heat Loss estimation mechanism

Estimating the amount of heat lost from a building based on infrared images is commonly done by calculating the thermal transmittance value, U-value, (Albaciti et al., 2010). However, to reliably calculate the thermal transmittance value, the difference in indoor and outdoor temperatures must be 10–15 ºC, the wind speed must be less than 1 m/s, the image must be captured during the winter, and emissivity must also be measured (Kim et al., 2019). As such, many photographing conditions, and additional measurements such as emissivity are required, it is difficult to calculate the thermal transmittance value in real time. In addition, recent research has proposed a method for automatically detecting abnormal temperature distribution (e.g., Asdrubali et al., 2018; Garrido et al., 2019; Park et al. 2021).

As the defective thermal image region has a bimodal temperature distribution, a methodology for determining the relative temperature threshold which determines the region with heat loss was used. In these studies, the bimodal distribution represents two combined normal distribution curves and two maximum peaks (Garrido et al., 2019). And the threshold is set to a value corresponding to the minimum value between the two maximum values (Asdrubali et al., 2018). In this study, the estimation of heat loss was carried out using previous studies that a bimodal distribution was formed in the presence of heat loss.

In addition, the method proposed in this study is to relatively compare the same area by dividing the area in the building envelope. The reason for classifying the same areas is that when measuring a thermal image, the radiation energy is affected by various factors such as the surface properties, temperature, wavelength, and emission direction of the material. It can be said that within a single image the factors are the same except for the surface properties. This is because the difference in the thermal pattern on the surface of the same material in the thermal image reflects the actual temperature difference (Park et al, 2021). Finally, in this study, heat loss is estimated through bimodal shaped temperature distribution, and the type of heat loss is provided.

Methods

The proposed process is divided into three steps, and the required input is only infrared images (Figure 3). The final output comes in two ways; first, information about the type, location, and proportion of objects presents in infrared images, and second, information about the area where heat loss is expected. This information can be used to determine the heat loss through the building envelope.
In this study, in heat loss estimation, comparisons are required for the same area. Therefore, comparison and analysis were conducted around the window among the components of the building envelope.

Datasets

![Simulation data](a) Simulation data

![Real building data](b) Real building data

*Figure 4: Example of Datasets*

The dataset for the train and test of Object Detection consists of an infrared simulation dataset, FAÇADE dataset, and our custom dataset, with a total of 1,425 images. With the absence of benchmark datasets reflecting the characteristics of infrared images in buildings, this study utilized infrared images of buildings through thermodynamic simulations. In addition, the custom dataset was developed by web-scraping to obtain images that reflected the features of modern buildings.

**Infrared Image producted Simulation:** In order to build a real heat loss environment and generate infrared images, simulation setup was very important to handle the overall heat transfer process from radiant heat transfer to convection heat transfer. The aforementioned actual environment was implemented by calculating the amount of radiation that varies from wavelength to characteristic of the air using MODTRAN tool to calculate the amount of radiation that varies from wavelength to characteristic. Reliable heat loss simulation image dataset was generated using the implemented environment and COMSOL tool. Additionally, building features used for simulation were 3D modelled based on actual building drawings, and for the best model to proceed with the simulation, interior walls inside buildings that did not affect the ambient heat loss were removed and simplified (Figure 4(a)).

**Real World building Image:** Real world image dataset was built with 378 images of Façade (Tylecek et al, 2013) and 567 images of custom building (Figure 4(b)). The FAÇADE image was constructed to train the regular and 567 images of Façade (Tylecek et al, 2013) was built with 378 images of Façade (Tylecek et al, 2013) with the two datasets corresponding to the Real Building Image dataset. The custom dataset was developed by web-scraping to obtain images that reflected the features of modern buildings.


Once in full-scale training, the difference between predicted result and ground truth is reduced based on loss function. The loss function used to train the model is as shown in Equation (1). The loss function has $L_{\text{class}}$ for location accuracy and $L_{\text{ann/final}}$ for class accuracy.

\[
L_{\text{CBU}} = 1 - bU + \frac{\rho^2 (b, b^\theta)}{c^2} + \alpha v \quad (2)
\]

\[
L_{\text{ann/final}} = \sum_{i=0}^{5} \sum_{j=0}^{B} \left[ - \log(p_i) + \text{BCE} (\hat{n}, n) \right] \quad (3)
\]

\[
L_{\text{classification}} = \sum_{i=0}^{5} \sum_{j=0}^{B} \left[ - \log(1 - p_i) \right] \quad (4)
\]

\[
\text{BCE} (\hat{n}, n) = -\hat{n} \log(n) - (1 - \hat{n}) \log(1 - n) \quad (5)
\]

\[
\alpha = \frac{v}{1 - bU + v} \quad (6)
\]

\[
v = \frac{4}{\pi^2} \left[ \text{atan} \frac{\omega_R^\theta}{\lambda} - \text{atan} \frac{\omega_R^\theta}{\lambda} \right] \quad (7)
\]

$L_{\text{CBU}}$ (Equation 2, 6, 7) is used to reduce the difference between the predicted box and the ground truth box.
\(d^2(b,b')\) is the Euclidean distance between the center point of the precipitation box and the ground truth, and \(c\) is the smallest distance of the area. \(\alpha\) is the positive trade-off parameter, \(\nu\) measures the consistency of aspect ratio.

\(L_{\text{confidence}}\) (Equation 3, 5) is divided into parts where and object exists and parts that do not exist so that it can be determined by placing more weight on the parts where the object exists. Where \(S^2\) represents the \(S \times S\) grid, each of which generates \(B\) bounding boxes over the network.

Finally, \(L_{\text{classification}}\) (Equation 4) uses cross entropy errors and it used to reduce differences in classification performance.

**Edge detection and Actual area calculation**

The purpose of this step is to find the ratio of the building to the actual window within the bounding box (i.e., as a result of an object detection in the form of a box) of the detected object in the entire image. To this end, we utilize object regions in the form of boxes, the output of object detection. We used the method of Canny edge detection in this step.

Then, Hough transform is applied to the image with edge detected. Especially, the Probability Hough line transform was used to find the coordinates of the straight lines in the results of the canny edge detection and to indicate where the object exists in the box. This process is based on the assumption that an object exists in its largest form with a straight line in the bounding box where it is found. Using the probability hough line transform, find the coordinates above the edge that are closest to the eight control points, the four vertices and the four centres of the edges, when the image is called rectangular, on the outside of the image (Figure 5 (b)).

**Figure 5:** (a) Edge detection result, (b) Example of finding a point above the edge closest to control points, (c) Example of calculating the actual area of an object.

The area in pixels of the object is subsequently obtained, such as Figure 5 (c). In this process, we use the method of taking one of the eight points as reference and obtaining the coordinates and triangular area of the other two consecutive points (Equation (8)).

\[
\sum_{i=2}^{7} S(A_i, A_i, A_{i+1})
\]

The function \(S(x, y, z)\) is a function that determines the area of the triangle formed by points \(x, y,\) and \(z,\) and \(A\) represents a set of eight points closest to control points. (\(i=2\ldots7\))

**Figure 6:** Image object detection & determining the actual area from detected edge

**Figure 7:** Heat loss estimation process

(The x-axis of the distribution means 279 classes of temperature, and the y-axis means the number of pixels corresponding to each class.)
As a result of edge detection and area calculation, it outputs the ratio of building to the image and windows to the building, such as Figure 6.

Heat loss estimation

It is the process of finding windows with relatively high temperature distributions by comparing temperature distributions with other windows present in a building, and through this, estimation by building envelope area is carried out. The detailed process can be seen in Figure 7.

In general, thermal meta data (such as radiant energy, radiation, reflective temperature, relative humidity, etc.) is required to calculate the temperature within infrared images (Garrido, I, 2019). However, in this work, we analyze relatively using only pixel values (RGB information) of infrared images without converting from infrared images to temperature information. This pixel values are classified by the distribution of the color spectrum. Each class consists of 279, which is matched to the degree of temperature. The class 1 is close to black and includes low temperature colors and class 279 includes colors that are close to white and red and have a high temperature.

To compare windows relatively, we first check the distribution over the entire window. Then, the class corresponding to the minimum point between the peak points shown in the window’s bimodal distribution is set as the threshold where heat loss occurs.

The estimation is carried out for each detected window using this defined threshold. Additionally, the type of heat loss is specified and provided, as follows.

- Type A: more than 20 percent of pixels above the threshold in one window
- Type B: 5-20 percent of pixels above the threshold in one window
- Type C: less than 5 percent of pixels above the threshold in one window

These categories were determined by analysing heat loss data for South Korean buildings. The heat loss threshold used in this process and the percentage of abnormalities in windows can be subjected to strict conditions, by lowering the threshold or lowering the abnormal judgment rate if users want more precise detection according to their needs.

Results

Object detection model training results

The neural network model for object detection is trained on 997 samples, which was 70% of 1,425 images structured for this study. Its performance evaluation of the trained model consists of 428 samples.

\[
\text{Precision} = \frac{TP}{TP + FP}, \quad \text{Recall} = \frac{TP}{TP + FN} \quad (9)
\]

\[
\text{IoU} = \frac{\text{area of Overlap}}{\text{area of Union}} \quad (10)
\]

For evaluation of the classification performance of the model, we used Precision(positive predictive rate) and Recall(true positive rate) in Equation (9).

Heat loss estimation results

A detected object region is evaluated according to Intersection over Union (IoU) value. In Figure 8, the green box is the area where the object detection model expected objects to exist, and the yellow box is the ground truth, the area where the actual object exists. IoU is the sum of the area of overlap between the two boxes, as shown in Equation (10).

- Precision = 0.96
- Recall = 0.97
- F1-score = 0.96
- Average IoU = 85.64%

The trained model performed sufficiently well for us in the proposed process. Its performance was compared with Windows Detector (Ali et al., 2007), Fast R-CNN, and SSD. The YOLO-V4 used in the proposed process had the highest performance with 97.31% AP and 98.59% Precision for window class.

<table>
<thead>
<tr>
<th>Model</th>
<th>Dataset</th>
<th>AP</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Window Detection</td>
<td>-</td>
<td>-</td>
<td>87.96%</td>
</tr>
<tr>
<td>Faster R-CNN</td>
<td>Building data</td>
<td>95.21%</td>
<td>96.32%</td>
</tr>
<tr>
<td>SSD</td>
<td>Building data</td>
<td>95.56%</td>
<td>96.98%</td>
</tr>
<tr>
<td>YOLO V4(Ours)</td>
<td>Building data</td>
<td>97.31%</td>
<td>98.59%</td>
</tr>
</tbody>
</table>

Table 1: Performance evaluation about Window class
building in Figure 9 due to the high temperature of the window where heat loss occurred. As such, the threshold is not selected by fixed criteria, but relatively determined by the context of infrared images.

Discussion and Future Research

Given that the model had a 96% F1-score for classification of classes and the 86% accuracy (average IoU) for location of object, we have achieved sufficient performance in extracting windows and other objects from images. The proposed process for the heat loss estimation was tested on two sample images. The input infrared image clearly showed heat loss. However, the degree of heat loss was classified as Type A/B/C state, resulting in numerical results.

Although this study focused on windows, the ultimate goal of this line of research is to derive metrics of heat loss for efficiently managing buildings. Therefore, additional research on the envelope of a building can lead to ideal results. Ultimately, this line of research is intended to split building envelope based on this study's proposed window-focused process. If we extract the envelope of the building except the window from the information of our works, we could get only envelope easily such as "the outer wall of the first room on the second floor". Applying the similar method used in this study, heat loss estimation can be carried out based on temperature distribution because the materials on the walls of buildings are made of the same material.

This line of research will be continued to develop heat loss estimation models for the building areas, including windows and exterior walls, to manage them more effectively.

Conclusion

This study proposed an end-to-end process using convolution neural networks that can estimate heat loss from buildings. This model consists of detecting objects in infrared images and estimating heat loss through them. Our datasets can be used as a benchmark by others to train neural networks to detect objects on the envelope of buildings.

In model, by changing this general image to grayscale, we create a model that detects infrared images well. The object detection model in the proposed process performed well after images were changed to grayscale. Objects' borders were accurately identified using edge detection. In estimation, based on comparison of relative temperature distributions, we determine the type of heat loss.

The proposed process does not require any user intervention other than inputting an infrared image. Users can also change the strictness of the estimation, which makes the proposed process flexible. This study is focused on developing end-to-end framework based on CNN to estimate heat loss. The proposed process can be utilized in real time, so it can be used to manage building energy efficiency.

Acknowledgement

This work was supported in part by the Korea Ministry of Land, Infrastructure and Transport (MOLIT) as [Innovative Talent Education Program for Smart City]
References


Figure 3: Copyright 2017. GUIDAFINESTAR Normativa, "Termica dei serramenti. Tutto quello che c’è da sapere all’Innovation Show", all rights reserved, https://www.guidafineastra.it/termica-dei-serramenti-tutto-che-ce-da-sapere-all-innovation-show/

Figure 9: Copyright 2017. GLASSFILM S.r.l. all rights reserved, https://www.glassfilm.it/component/k2/dispersione-termica-vetrato

Figure 10: Copyright 2021. Open Eco Homes, all rights reserved, http://openecohomes.org/equipment-articles-and-advice/