energies

Article

A Novel Building Temperature Simulation Approach Driven by Expanding Semantic Segmentation Training Datasets with Synthetic Aerial Thermal Images

Yu Hou [1,*], Rebekka Volk [2] and Lucio Soibelman [1]

[1] Sonny Astani Department of Civil and Environmental Engineering, University of Southern California, Los Angeles, CA 90089-1453, USA; soibelman@usc.edu

[2] Institute for Industrial Production, Karlsruhe Institute of Technology, 76131 Karlsruhe, Germany; rebekka.volk@kit.edu

* Correspondence: yuhou@usc.edu

Abstract: Multi-sensor imagery data has been used by researchers for the image semantic segmentation of buildings and outdoor scenes. Due to multi-sensor data hunger, researchers have implemented many simulation approaches to create synthetic datasets, and they have also synthesized thermal images because such thermal information can potentially improve segmentation accuracy. However, current approaches are mostly based on the laws of physics and are limited to geometric models' level of detail (LOD), which describes the overall planning or modeling state. Another issue in current physics-based approaches is that thermal images cannot be aligned to RGB images because the configurations of a virtual camera used for rendering thermal images are difficult to synchronize with the configurations of a real camera used for capturing RGB images, which is important for segmentation. In this study, we propose an image translation approach to directly convert RGB images to simulated thermal images for expanding segmentation datasets. We aim to investigate the benefits of using an image translation approach for generating synthetic aerial thermal images and compare those approaches with physics-based approaches. Our datasets for generating thermal images are from a city center and a university campus in Karlsruhe, Germany. We found that using the generating model established by the city center to generate thermal images for campus datasets performed better than using the latter to generate thermal images for the former. We also found that using a generating model established by one building style to generate thermal images for datasets with the same building styles performed well. Therefore, we suggest using training datasets with richer and more diverse building architectural information, more complex envelope structures, and similar building styles to testing datasets for an image translation approach.

Keywords: building envelopes; thermal image simulation; segmentation datasets; data hunger

check for updates

Citation: Hou, Y.; Volk, R.; Soibelman, L. A Novel Building Temperature Simulation Approach Driven by Expanding Semantic Segmentation Training Datasets with Synthetic Aerial Thermal Images. *Energies* **2021**, *14*, 353. https://doi.org/10.3390/en14020353

Received: 28 November 2020
Accepted: 8 January 2021
Published: 11 January 2021

1. Introduction

Unmanned aircraft systems (UASs), also known as drones, have commonly been used in civil engineering and military applications [1]. For example, UAS-based aerial images integrated with photogrammetric technologies allow for classifying building elements [2], monitoring and controlling construction sites [3], and creating virtual environments for mission planning and rehearsals [1]. The photogrammetric technology, which maps images acquired by drones onto a 3D model, provides simple analytics, for example, distance and dimension measurement. Integrated with other tools and applications, a photogrammetry-recreated 3D model can detect not only structural damage but also heat loss from buildings and district heating networks. Such a 3D model can also be used to locate roads and classify their materials to precisely calculate driving time for route planning. All these examples emphasize on the need for extracting semantic information from photogrammetry-recreated models.

To extract semantic information, also known as semantic segmentation, from images or photogrammetric models, many computer vision algorithms, especially deep learning approaches, have been applied, such as MaskRCNN [4], Yolo family [5], and DeepLab family [6]. Early studies used images or 3D models with only one channel (mostly the red-green-blue (RGB) color channel obtained by an image sensor). However, segmentation based on single sensor images is insufficient when facing complex scenarios; thus, for more accurate classification and segmentation, researchers have added more channels and features to RGB images [7]. For example, Chen et al. [1] added texture, point density, local surface, and open source features, while Liu et al. [8] added depth information to improve photogrammetric point cloud segmentation. Researchers have also improved segmentation by adding thermal information [7,9].

Despite the great success achieved by the previously described studies, deep learning algorithms are quite data hungry as demonstrated in many studies [10,11]. Data hunger refers to the size of the training dataset required for generating a model with a good predictive accuracy [12]. It is difficult for individual research groups to expand the training datasets because researchers are often unwilling to share data, or their data formats are incompatible. Therefore, researchers are forced to collect more data on their own. However, collecting data usually takes several days for a large district and is labor-intensive, costly, and inefficient [13]. Additionally, annotating these new acquired training datasets also requires many hours of labor and inspection for annotation accuracy. In order to solve the data hunger problem, some researchers have used synthetic data. For example, Chen et al. [14] designed a framework to generate synthetic images from a 3D virtual environment. They simulated drone flight paths over the synthetic virtual environment (3D point cloud) that had annotated information such as the ground, buildings, and trees to render synthetic images with corresponding annotations. In their framework, depth images, which can be obtained by Lidar, and RGB images, which can be obtained by color cameras, in the real world were instead generated virtually. Data hunger also occurs with images that fuse RGB with thermal information. For example, Li et al. [15] used thermal images taken outdoors and indoors on the ground to segment pedestrians, cars, tables, lamps, and other objects. This takes advantage of the thermal camera's ability to capture information in dark and hazy environments. They also introduced synthetic thermal images to improve segmentation. Inspired by Li's studies, we planned to use thermal information to improve the segmentation of aerial images of buildings outdoor scenes because it would allow us to capture different thermal signatures of each part of the building and its surroundings [10,16,17]. Segmentation of building components has many benefits for energy analysis such as detecting building envelopes' heat loss, moisture, and thermal bridges [18–20]. It also allows for simulating energy consumption [21,22]. Therefore, generating synthetic thermal images as complementary information for RGB images could further improve the segmentation process.

There are several simulation approaches to generate synthetic thermal information. For instance, physics-based building surface thermal simulation enables the precise quantification of energy fluxes and simulates the building surface temperatures by using heat equations [23–25]. Many recent studies have used 3D models to simulate heat transfers [26,27]; however, these studies are limited to their level of detail (LOD), and accuracy and effectiveness are reduced [23,28–30]. To be precise, there is no surface temperature simulation based on an as-is built model (the highest LOD model) due to the computational complexity and inherent uncertainties caused by the many default parameters and assumptions used in a simulation process [31]. Furthermore, physics-based simulation mainly works for buildings or specific infrastructures not for the surrounding environment such as trees or streets. Physics-based simulation methods do not simulate the surrounding environment in detail and simplify the surrounding environment as boxes in the geometric models [25].

Unlike the aforementioned approaches, this study focuses on simulating temperature information for generating synthetic aerial thermal images. Our approach learns features and extracts information from historical data of drone-based images instead of a physics-based thermal simulation. Our approach avoids using default configurations when detailed system information such as building material and users' behaviors is not available. Furthermore, our approach is not limited to geometric models' LOD; on the contrary, current approaches depend on 3D models' precision for accuracy. Our approach implements computer vision algorithms to translate RGB images acquired by drones over a large-scale area to thermal images, which also enables them to be fused with RGB images for segmentation with multi-sensor data. Our study is designed to answer the following questions: (1) How can RGB images of buildings be used to generate thermal images? (2) How can training data of captured RGB images affect simulation results? Particularly, how is the generation model established by one building style used to generate thermal images with another? (3) What are the similarities and differences between the current approaches and our proposed approach for generating thermal images? This study will only focus on thermal image generation performance by evaluating the generated results compared to the ground truth. The performance of deep learning using generated images will be evaluated in a future study. The rest of this paper is organized into the following sections: Sections 2 and 3 review the current work of surface temperature simulation and computer vision techniques that have been used in this study. Section 4 presents the methodology of this study. Section 5 presents results and discussion. Section 6 concludes the paper and presents future work possibilities.

2. Thermal Simulation for Building Envelope to Generate Thermal Images

A building model's complexity affects the resolution of the surface temperature simulation for generating synthetic thermal imagery data [32]. Some case studies have shown that less comprehensive models only simulate one unified surface temperature per whole façade, while more complex models can incorporate more specific parameters of facades such as their material, orientations, and functions, which makes the temperature simulations more accurate [25,29,32]. For example, Aguerre et al. [33] designed and implemented ThRend, a facade thermogram simulation tool, which rendered building thermal images based on different components' emissivity and reflectivity configurations [25]. The models used in their experiments were simplified down to four uniformly thick boxes representing street buildings. Therefore, the simulation results could not capture slight temperature changes on the facades. For example, at 4 a.m. and 7 p.m., the facades in the simulated thermal images have uniform simulated thermal temperature. Additionally, the results cannot simulate thermal bridges on the walls. In their new studies that were developed as an incremental improvement over their previous work [34,35], they integrated a higher level of detail geometry into a finite element method (FEM) solver. Their simulation results were more accurate and detailed, for instance, their results could simulate the temperature changes between windows and walls.

There are other similar studies conducted by Henon et al. [28,36], Kottler et al. [23,24,29], and Xiong et al. [30,37]. Henon et al. conducted their experiments using software SOLENE, which can simulate the climatological factors of urban neighborhoods, to compute building surface temperature and evaluate the sensible heat flux to the city atmosphere [28,36]. Kottler et al. chose a physical approach to simulate the building surface temperatures using heat equations [23,24,29]. First, different building components simulated in the models were clearly classified and linked to a material library. Second, the vegetation and trees were also taken into consideration when the building surface temperature was simulated, but the model in their experiment was still simplified. As the authors described in their research, vegetation and trees were roughly represented as so called forest boxes. Furthermore, trees that were close to each other were integrated in one model, and single trees were ignored. Xiong et al. argued that geometric model generation for simulation was labor-intensive and time-consuming; therefore, they implemented a method to semi-automatically simulate temperature signatures [30,37]. Their simulation pipeline included 3D model reconstruction, component classification, model surface subdivision, material assignment, and infrared rendering with limitations due to the level of detail (LOD). LOD is a term describing the overall planning or modeling state at a certain time for design and construction. It can present the complexity of 3D model visualization [38]. Xiong et al.'s simulation was based on an approximated mesh model by using planar primitives. Their simulation model was classified as LOD 2, at which roofs and facades were illustrated, but detailed objects of the roofs and facades were not generated. Therefore, synthetic thermal imagery data generated by such approaches do not allow the growth of semantic segmentation training datasets.

Current simulation approaches, based on physics equations and laws of thermodynamics, are limited to model details and computational time. These approaches are known as deterministic systems or a "white box." However, many stochastic systems or "black box" approaches like neural networks have achieved competitive results. It is then necessary to study stochastic systems that allow to simulate objects' temperatures for the generation of synthetic data.

3. Computer Vision and Generative Adversarial Networks (GAN)

As reviewed in the last section, earlier temperature simulation tasks have used mathematical and physics models to predict the energy transfer between indoors and outdoors to estimate the building envelope temperature. However, such approaches bring with them many assumptions and can be limited due to the artificial models used for simulation. Therefore, researchers have been attempting to improve models' LOD to the highest as-built level by deploying computer vision and photogrammetry since RGB images can directly record the buildings' appearance, which inspires researchers to interpret the information behind the images, such as image-to-image translation that converts RGB images to thermal images [39,40]. There are many computer vision approaches used for image translation, and the most robust and successful approach is generative adversarial network (GAN). The approach we propose to use in this study is also based on GAN.

3.1. Computer Vision and Neural Networks

Computer vision tasks include collecting, processing, extracting, analyzing, and understanding digital images. There are many computer vision applications in civil engineering [41], such as damage detection [42–45], change detection [46–49], and structural component recognition [50–53]. Additionally, computer vision is also used in urban energy tasks such as detecting leakage for district heating networks [18,54,55], identifying thermal bridge and moisture on building envelopes [19,20], and simulating energy consumption based on thermal images [21,22].

Compared to more traditional computer vision approaches, convolutional neural networks (CNNs) introduce non-linearity, which considers the dependency between each pixel in an image. Krizhevsky et al.'s AlexNet is considered the pioneer use of CNNs [56]. After them, VGGNet [57], ZFNet [58], and GoogleNet [59] had improved image processing performance. These computer vision approaches have been used for object recognition, semantic segmentation, scene reconstruction, and many other topics. Researchers have also discovered the potential of computer vision in translating and generating images.

3.2. From "Unstructured" to Conditional Generative Adversarial Network (GAN)

To translate and generate images, researchers have adapted structures and neurons of hidden layers in the original CNNs. Since image-to-image translation tasks are pixel-wise classification and regression based [39,60–62], researchers need to modify the output layers to generate images. In earlier work, the formulations and processes usually transformed input to output in a "unistructural" way, which means that each pixel is independent of other pixels. Nevertheless, conditional generative adversarial networks (GANs) proposed by Goodfellow et al. [63] instead learned a "structured loss" that considered the joint features of the output pixels. Additionally, loss function, an important technique, is different from other network approaches in conditional GANs [6,64,65]. Conditional

GAN is a machine learning framework used to generate information such as a block of text or a robot's action. It is formed by the generative network which can generate candidates (sentences or actions) and the discriminative network which evaluates the generated candidates. Due to the rapid development of neural networks, the performance of generating information using GAN has been improved in recent years. Not only can a GAN generate sentences and actions, but it also has applications in image translation.

GAN has been used for many image-to-image transformation applications including image prediction (next frame prediction [66], product photo generation [67]), image generation from sparse annotations [68,69], and painting style transfer [70]. However, these approaches are application specific. Isola et al. [39] and Zhu et al. [40] proposed a "pixel2pixel" and a "cycle-consistent" GAN. Their approaches are not task-specific. Particularly, Zhu et al.'s work also learned the input-output image mapping without the need for paired training examples.

4. Research Method

4.1. Research Design

This study includes three steps: (1) dataset preparation, (2) building envelope thermal image rendering, and (3) evaluation. Figure 1 illustrates the research method workflow.

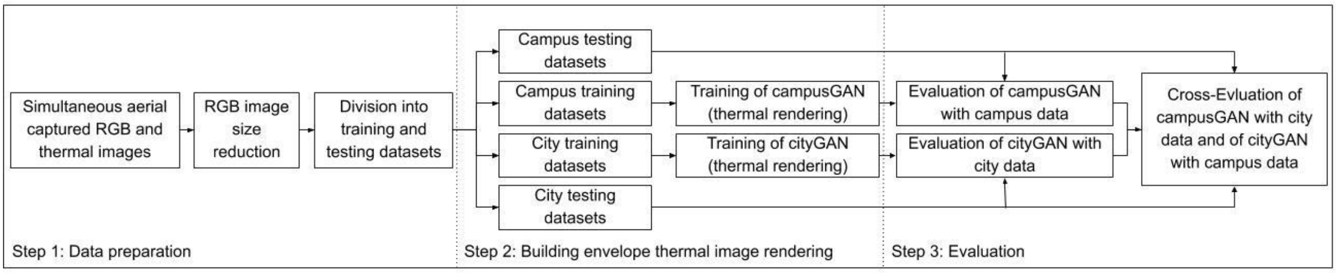

Simultaneous aerial captured RGB and thermal images
RGB image size reduction
Division into training and testing datasets
Campus testing datasets
Campus training datasets
City training datasets
City testing datasets
Training of campusGAN (thermal rendering)
Training of cityGAN (thermal rendering)
Evaluation of campusGAN with campus data
Evaluation of cityGAN with city data
Cross-Evluation of campusGAN with city data and of cityGAN with campus data
Step 1: Data preparation
Step 2: Building envelope thermal image rendering
Step 3: Evaluation

Figure 1. Research method flowchart.

The first step is data preparation. In order to fit the algorithm and save memory and computation time, the resolution of aerial captured RGB images needs to be reduced. Additionally, the method used in this approach requires the whole dataset to be divided into training and testing datasets following a commonly used proportion. In this study, our datasets included both campuses and city areas.

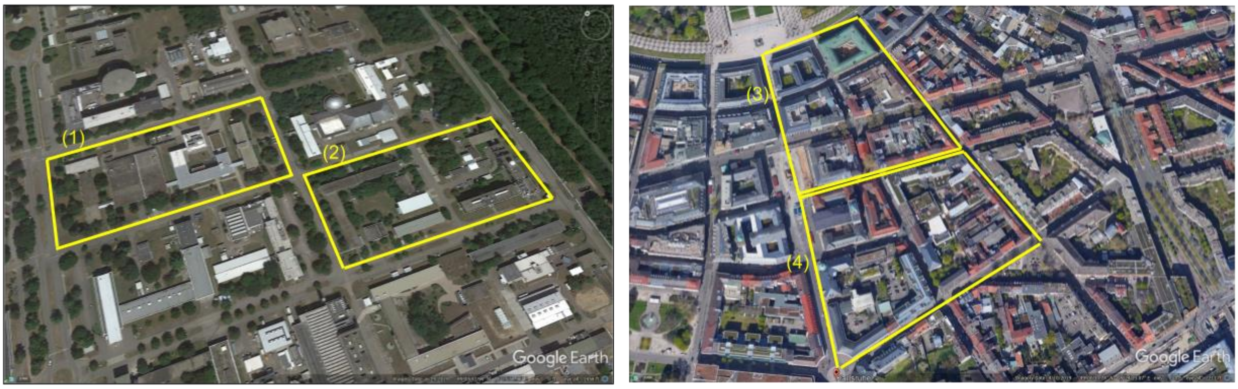
The second step is rendering building envelope thermal images. The image translation neural network was introduced in this step, and its network parameters were trained and updated by campus and city training datasets. Such trained network models were used to simulate thermal images.

The last step is evaluation of the proposed method. The simulated thermal images were compared with ground truth via two evaluation criteria including two mathematical approaches: pixel-wise mean squared error (MSE) and structural similarity index (SSIM). The evaluation criteria were conducted on campus and city data with their respective testing data as well as cross evaluation.

4.2. Simulation Domains and Dataset Preparation

To easily detect the thermal image contrast of building envelopes, we collected data in the winter in Karlsruhe, Germany, since the temperature difference between indoors and outdoors is obvious there in the winter. The experiments were conducted in two structurally different outdoor scenes. One was on a college campus where modern buildings, separated by lawn and roads, were not close in proximity in a suburban area. The other one was in a dense city area in Germany where traditional multi-story European buildings were located close together. The reasons for the selection of these two scenarios are that (1) the heat island effect is more obvious in city areas than in suburban areas; and (2) architectural styles of buildings are different in city areas and in suburban areas. Conducting experiments in both areas allowed us to comprehensively explore our approaches.

In this study, we designed four experiments. Thermal and corresponding RGB images were taken from two separated areas on the campus area for experiments one and two, as shown in Figure 2a. These two experiments are abbreviated as "Camp1" and "Camp2". Images were also taken from two separated areas in the city area for experiments three and four, as shown in Figure 2b and abbreviated as "City1" and "City2".



(**a**) Two experiments on campus.

(**b**) Two experiments in a European city center

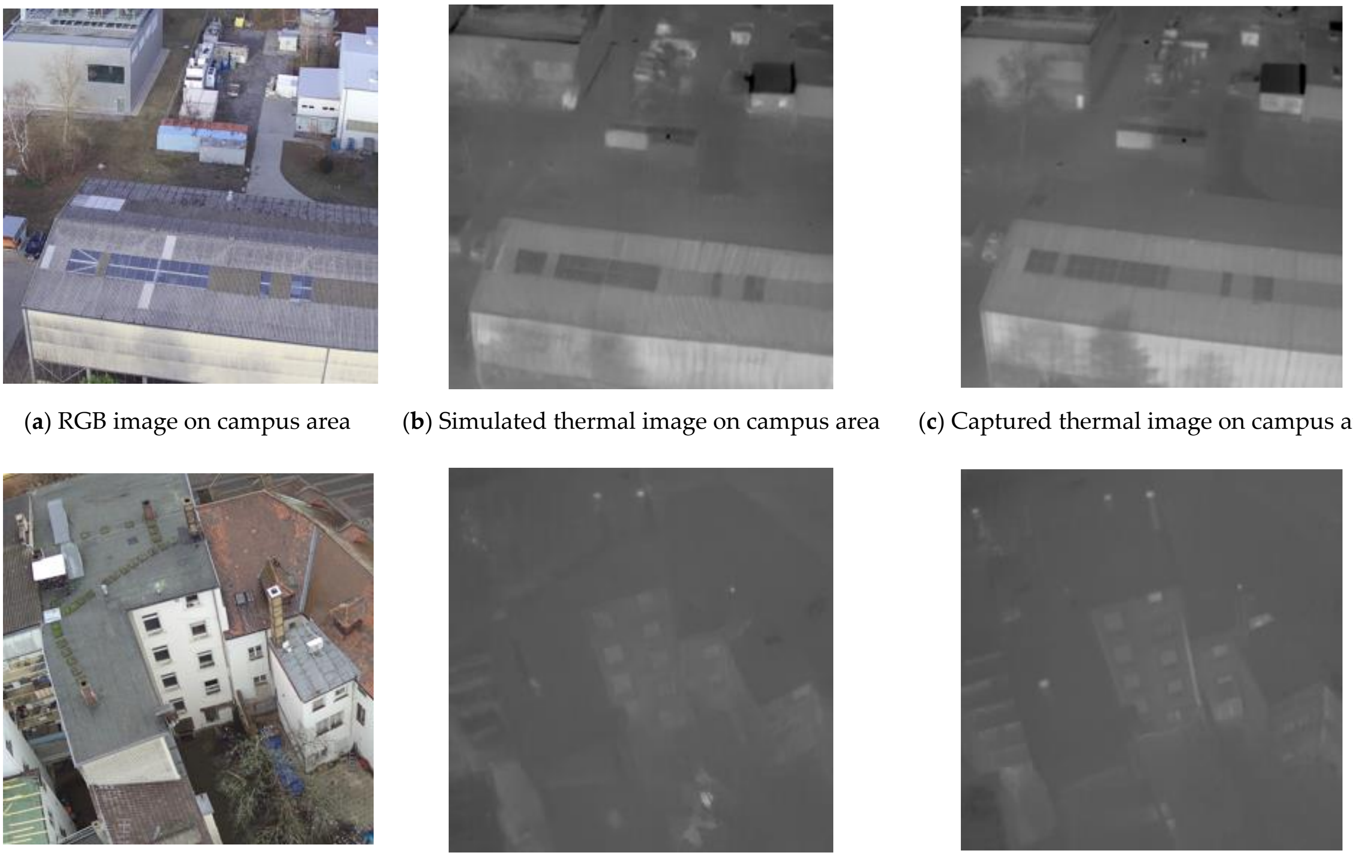
Figure 2. Illustration of the experiment locations.

In order to keep the size of the dataset balanced, there were around 20,000 images in each experiment. Each image had a resolution of 2048 × 2048 pixels^2 and was resized to 256 × 256 pixels^2 to save algorithm computing memory and time. It is common to divide datasets into training and testing sets. Training dataset usually accounts for 70% (14,000 images) of the whole dataset, and testing dataset accounts for the rest. Therefore, datasets in these four experiments were all divided into training and testing datasets.

4.3. As-Built Building Envelope Thermal Image Rendering

The algorithms used in this paper were based on Isola et al.'s previous work called "pixel2pixel", which is an image-to-image translation based on GAN and is not task specific. The network architecture used inside of the algorithm is a fully convolutional network (U-net). The basic theory of image translation used in this study is to directly convert RGB images to thermal images via networks.

In a commonly used neural network flowchart, training datasets can train and build a learning model. The learning model learns rules and features from training datasets by comparing predicted results with ground truth. After the model learns features by continuously adjusting its parameters, it can process testing datasets with its updated parameters. In this study, datasets consisted of pairs of real captured RGB and thermal images by cameras. The RGB images in proportionally separated training datasets were fed into the initial GAN model. The model then converted RGB images to simulated thermal images, and updated its inner parameters based on reducing the discrepancies between simulated thermal images and captured thermal images to improve the simulation performance. After many rounds of updating parameters (200 epochs in this study), the GAN model was ready to process RGB images in proportionally separated testing datasets. Since we had four experiments as shown in Figure 2, we had four training datasets and four testing datasets. In this study, we used a training dataset and built a GAN model to process a testing dataset not only in the same experiment but also in a different experiment. The cross-evaluation between every two experiments allowed us to observe how the generation model established in one building style could be used to generate thermal images on another. As the examples show in Figure 3, the GAN model converts the RGB images, Figure 3a,d, to the simulated thermal images Figure 3b,e. The latter images are compared with the captured thermal images in Figure 3c,f. Training datasets and testing datasets in Figure 3 are from the same experiment, so simulated and captured thermal images look identical. Other cases with cross-evaluations are illustrated and discussed in the results section, and some discrepancies between simulated and captured thermal images are observed. The shades of gray color in the thermal images indicate hotter areas (light gray) and colder areas (dark gray). We selected the black-white palette for two reasons. First, the monotonous color palette can intuitively represent the contrast between hot areas and cold areas. Second, the black-white palette only uses one channel to represent images, and we can color code from 0 to 255 to represent temperature information, which is easy for algorithms to calculate in the GAN model.



(**a**) RGB image on campus area (**b**) Simulated thermal image on campus area (**c**) Captured thermal image on campus area

(**d**) RGB image in city area (**e**) Simulated thermal image in city area (**f**) Captured thermal image in city area

Figure 3. Examples that explain thermal image rendering.

4.4. Evaluation Metrics

The performance of each experiment and cross-evaluation between every two experiments were measured by comparing rendered thermal images (R) generated from RGB images by the GAN model with real captured thermal images (C) by using image similarity evaluation criteria, pixel-wise mean squared error (MSE), and structural similarity index (SSIM), as shown in Equations (1) and (2) [71]. These criteria calculate pixel-wise per image and compare rendered with captured thermal images.

$$\mathrm{MSE}(R,\ C) = \frac{1}{mn}\sum_{x=0}^{m-1}\sum_{y=0}^{n-1}[R(x,y) - C(x,y)]^2 \tag{1}$$

$$\mathrm{SSIM}(R,C) = \frac{(2\mu_R\mu_C + c_1)(2\sigma_{RC} + c_2)}{(\mu_R^2 + \mu_C^2 + c_1)(\sigma_R^2 + \sigma_C^2 + c_2)} \tag{2}$$

In Equation (1), R represents a rendered image and C represents a captured image. The resolutions of those two images are both 256 pixels times 256 pixels. The character $(x,\ y)$ represents the same coordinate of pixels in both rendered and thermal images. The differences of every two relevant pixels in two compared images are evaluated by squaring these differences, summing them up, and dividing the sum of squares by the total number of pixels (256 × 256) in the images. An MSE of value 0 shows that two compared images are completely identical, and an MSE that is bigger than 0 indicates that two compared images are different. The bigger the MSE values are, the more differences the two compared images have, which means that the generation model renders a rendered image with more errors compared to a captured image. However, MSE is unable to agree with human subjective analysis [72]. Therefore, SSIM was selected as a complimentary evaluation approach.

SSIM is used to compare the structural information of images. In Equation (2), R represents a rendered image and C represents a captured image. Symbols μ_R and σ_R represent the mean value and standard deviation value of pixels in a rendered image, as shown in Equations (3) and (4), and μ_C, σ_C represent these values for a captured image. Symbol σ_{RC} represents the covariance of rendered images and captured images, as shown in Equation (5). Coordinate $(x,\ y)$ indicates the same coordinate of pixels in the compared two images. Last, symbols c_1 and c_2, in Equation (2) are constants used for the stability of the equation when μ and σ are extremely small. The range of SSIM value is between -1 and 1, where 1 represents perfect identicality.

$$\mu_R = \frac{1}{mn}\sum_{x=0}^{m-1}\sum_{y=0}^{n-1} R(x,y) \tag{3}$$

$$\sigma_R = \sqrt{\frac{1}{mn-1}\sum_{x=0}^{m-1}\sum_{y=0}^{n-1}(R(x,y)-\mu_R)^2} \tag{4}$$

$$\sigma_{RC} = \sqrt{\frac{1}{mn-1}\sum_{x=0}^{m-1}\sum_{y=0}^{n-1}(R(x,y)-\mu_R)(C(x,y)-\mu_c)} \tag{5}$$

5. Results and Discussion

There were four experiments used in this study, abbreviated as “Camp1”, “Camp2”, “City1”, and “City2”. The evaluations were conducted on the testing datasets in the same experiment and between different experiments as shown in Table 1. Each row represents a GAN model that was built based on a training dataset in a corresponding experiment, and this GAN model processes a testing dataset in each column. The color in Table 1 represents the value of the number in a cell. According to the evaluation metrics, higher MSE values or lower SSIM values represent worse performance. Therefore, the red color represents higher MSE and lower SSIM values, namely worse performance. To represent better performance, the green color represents lower MSE and higher SSIM values, and the red color represents higher MSE and lower SSIM.

To investigate the bad performances, we selected cases with the highest MSE and lowest SSIM values in each evaluation both in the same experiment and in the cross-evaluation, as shown in Figure 4. In the same way that the horizontal and vertical headers are organized in Table 1, each row in Figure 4 represents what training dataset was used to build a GAN model, and each column represents a testing dataset that such a GAN model processes. The titles “Real captured”, “Simulated”, and “RGB” in Figure 4 represent captured thermal images, rendered thermal images by a GAN model, and corresponding RGB images. The selected images have highest MSE and lowest SSIM in each evaluation.

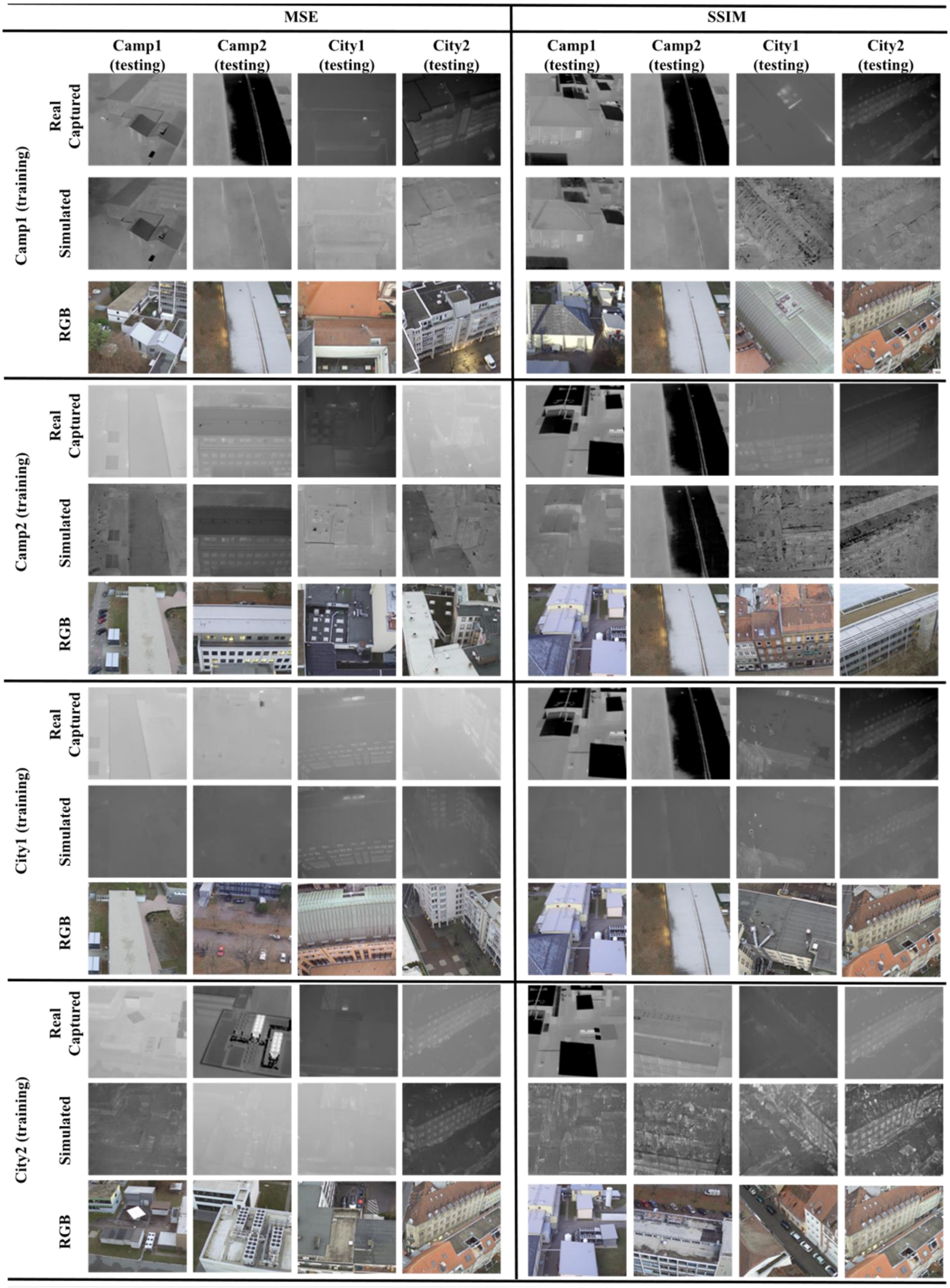


Figure 4. Selected images with highest MSE and lowest SSIM in each evaluation.

Table 1. Total average mean squared error (MSE) and structural similarity index (SSIM) values in each experiment.

	Total Average MSE				Total Average SSIM			
	Camp1 (Testing)	**Camp2 (Testing)**	**City1 (Testing)**	**City2 (Testing)**	**Camp1 (Testing)**	**Camp2 (Testing)**	**City1 (Testing)**	**City2 (Testing)**
Camp1 (training)	2.5779308	60.38710	81.72707	132.63658	0.927918	0.809343	0.803144	0.789476
Camp2 (training)	56.352844	4.675767	60.85602	138.23941	0.823244	0.914039	0.788834	0.75134
City1 (training)	107.46189	72.40453	3.70587	159.21241	0.837777	0.838546	0.944457	0.885718
City2 (training)	79.49741	66.63147	88.94927	2.33137	0.803554	0.800675	0.834157	0.943066

5.1. Simulation Result Assessment

As described in the method Section 4, we evaluated the GAN simulation approaches based on MSE and SSIM values. As shown in Table 1, if we see MSE and SSIM evaluations as two matrices, the color patterns for MSE and SSIM evaluation matrices are basically similar. First, green color is in the diagonals both for total average MSE and SSIM evaluations, in other words, we can observe a good performance in a case in which training and testing datasets both stem from the same experiment. Second, using a GAN model that is built based on city training datasets to render campus testing datasets performs better than the inverse, since values in upper triangular entries are higher than values in lower triangular entries in a MSE matrix, and values in upper triangular entries are lower than values in lower triangular entries in a SSIM matrix. The potential explanation is that the building styles in city centers are more complex than on campuses, which allows a GAN model to learn more hidden features from building envelopes in a city center. Additionally, as Figure 2a shows, buildings are sparsely located and separated by lawn and roads on campus. Thus, there is less building envelope information for a GAN model to learn. Therefore, a GAN model established by city datasets is more capable of simulating building envelope thermal information. Third, although the color patterns are similar between MSE and SSIM evaluation matrices, there is an outlier in an entry (city1 training dataset—city 2 testing dataset) in MSE matrix, which is supposed to be small, but such entry in SSIM evaluation matrix is normal.

Figure 4 illustrated the selected cases with bad performance in terms of MSE and SSIM values. As the MSE and SSIM metrics described, the simulated thermal images with highest MSE values have big color differences (grayscale color represents temperature information) from real captured thermal images, and the simulated images with lowest SSIM values have more image noise and difficulties in representing building envelope structures. Campus buildings' envelopes are not complex like city buildings' envelopes; therefore, campus testing datasets intuitively are simulated better than city testing datasets, although simulated images shown in Figure 4 are cases with highest MSE and lowest SSIM. Additionally, the observation that GAN models built based on city datasets perform better is also validated in Figure 4.

In order to understand the relationship between MSE and SSIM in terms of all images in an individual evaluation, we plotted several multivariate distribution figures for both the same and cross-experiment evaluations. In the same way that the headers are organized in Table 1, these distribution figures are plotted in Figure 5. In each distribution figure, the x-axis represents MSE value, and the y-axis represents SSIM value. Each image in the testing dataset has a MSE and a SSIM value, and a red point with a pair of MSE-SSIM coordinates is drawn in the distribution. The darker red area represents concentrated red points while the lighter red area represents scattered red points. There are some patterns observed in Figure 5. First, most figures illustrate negative correlations between MSE and SSIM values. Figures in the diagonal from upper left to lower right show robust negative correlations that red areas are very thin like a line with a strong negative coefficient. However, red points in other figures are scattered, which means such a performance is not stable. Second, if we observe the distributions in terms of MSE and SSIM values, respectively, we find that MSE values follow a long-tail distribution while SSIM values follow a Gaussian distribution in most evaluations. Third, distributions for the cases using a GAN model built with the city training dataset to process campus testing dataset are more stable than the cases in an inverse way. The reason is that distributions in Figure 5i,j,m,n are more stable than distributions in Figure 5c,d,g,h.

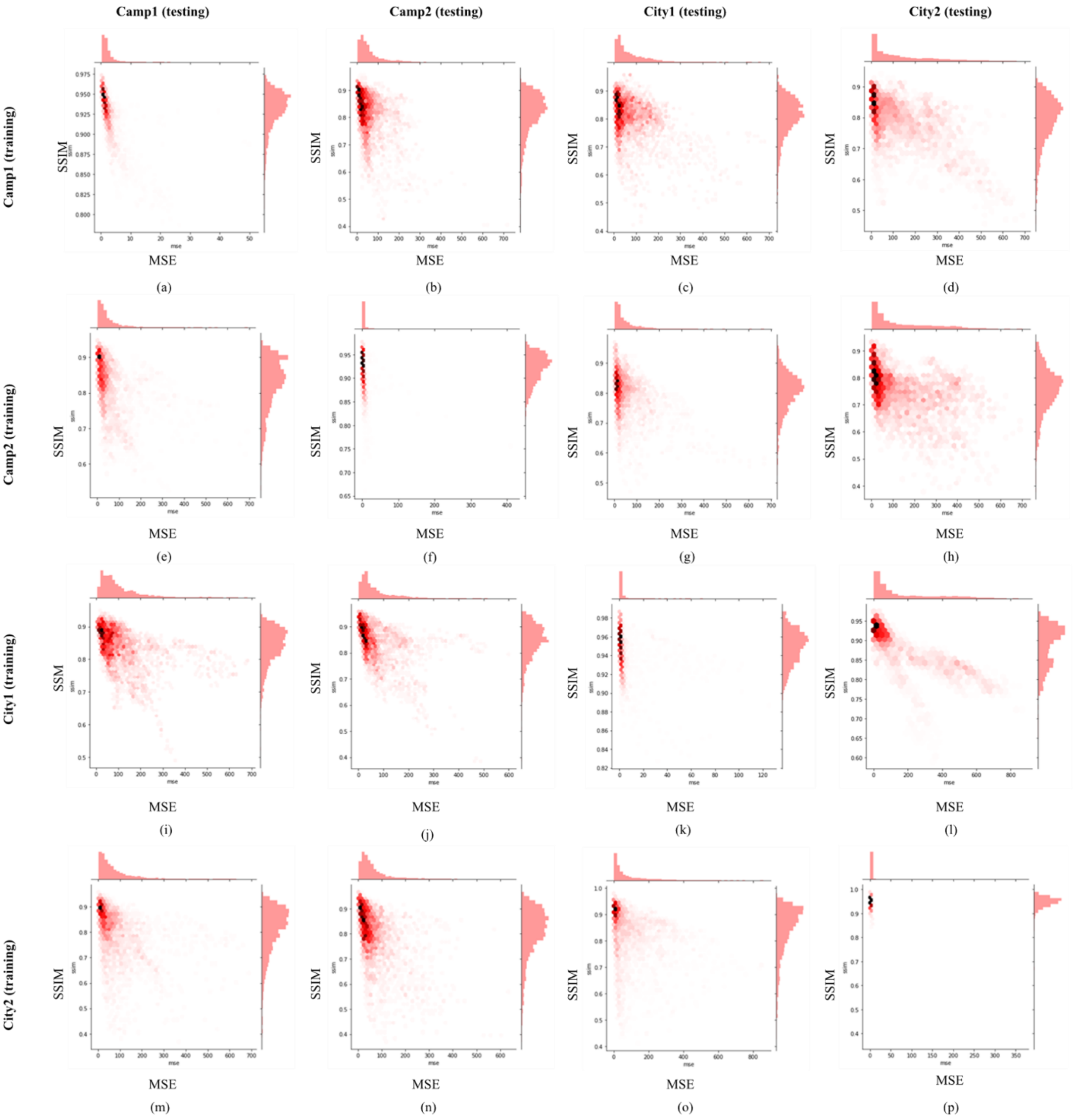


Figure 5. Multivariate distribution figures for both the same and cross-experiment evaluations. (**a**) Distribution of Camp1(training) vs. Camp1(testing), (**b**) Distribution of Camp1(training) vs. Camp2(testing), (**c**) Distribution of Camp1(training) vs. City1(testing), (**d**) Distribution of Camp1(training) vs. City2(testing), (**e**) Distribution of Camp2(training) vs. Camp1(testing), (**f**) Distribution of Camp2(training) vs. Camp2(testing), (**g**) Distribution of Camp2(training) vs. City1(testing), (**h**) Distribution of Camp2(training) vs. City2(testing), (**i**) Distribution of City1(training) vs. Camp1(testing), (**j**) Distribution of City1(training) vs. Camp2(testing), (**k**) Distribution of City1(training) vs. City1(testing), (**l**) Distribution of City1(training) vs. City2(testing), (**m**) Distribution of City2(training) vs. Camp1(testing), (**n**) Distribution of City2(training) vs. Camp2(testing), (**o**) Distribution of City2(training) vs. City1(testing), (**p**) Distribution of City2(training) vs. City2(testing).

5.2. Comparison between Our Results and Other Existing Methods

There have been several simulation tools to generate synthetic thermal images for growing deep learning training datasets. Our approaches have some similarities compared to the existing methods. First, we all can simulate the thermal information of building envelopes without limitations to the building styles. As the results showed, we simulated thermal images of building envelopes both on campuses and in city areas. Meanwhile, the existing methods also do not have difficulties in simulating building envelopes with different building styles. Second, we all can simulate thermal images for generating synthetic thermal images to some extent [15,73]. The existing methods need to simulate the 3D geometric model first, but the thermal images still can be rendered by a virtual camera in the 3D model. Our own approach has several differences from the current approaches. First, as Henon et al. [36] described in their approach, they omitted some small structures (appliances and chimneys) on roofs. In our study, there are many European traditional city buildings with appliances and chimneys on complex roofs. Since our approach is directly implemented on captured images, these features are not omitted. Second, the evaluation metrics are different. For example, in Aguerre et al.'s [34] experiment, their simulations were based on building models, as such their evaluation did not include the surrounding environment such as trees or streets. In contrast, our approach covered both buildings and their surroundings. In addition, Aguerre et al. compared simulation results from selected areas of building envelopes with real thermal information. Such comparison cannot cover areas that the building model did not represent in the simulation. Their evaluation failed to compare this issue. In our approach, we compared the simulated thermal errors by evaluating MSE values, on top of which, we also compared the building envelope structures in the simulated images by evaluating SSIM values. We observe that an image translation approach is more feasible than a physics-based approach for generating synthetic thermal images for segmentation datasets. Third, if a physics-based approach is used for generating thermal images, researchers should configure a virtual camera that is consistent to a camera used for capturing RGB images in terms of camera position, focal length, and point of view (POV), but such a virtual camera is difficult to accurately configure. Our image translation approach avoids these procedures because it directly converts RGB images to thermal images.

On the other hand, our approach also has drawbacks compared to current approaches. As Aguerre et al. [34] described, they can simulate the surface temperatures at different times of the day by adjusting parameters. However, our datasets were and should be captured during the same time span of the day. For example, datasets captured in the morning are not capable to simulate envelope surface temperatures in the night.

6. Conclusions and Future Work

Thermal information can be used to improve the segmentation of aerial images of outdoor scenes. We proposed an innovative image translation approach that would simulate temperature information and we analyzed and validated that such an approach is more feasible than a physics-based approach for generating synthetic thermal images for segmentation. Compared to current approaches, these are the main benefits to our approach: (1) It avoids acquisition of detailed system information like building materials and does not require default configurations. This is more feasible for old buildings that lack detailed information. (2) Our approach is not limited to the geometric models' precision and LOD, since image data used in our approach are taken from drone view directly capturing the as-built building envelopes. Our approach can save time compared with creating a geometric digital model. (3) Our approach can simulate buildings' surrounding environment thermal information such as trees and streets. Those elements were simplified in physics-based approaches as boxes during simulation. (4) Since our approach directly converts RGB images to thermal images, it does not need to align a virtual camera that renders thermal images to a real camera that captures RGB images.

Our approach also has some limitations. Since the simulation process is based on historical training datasets instead of the laws of physics, the time and season when these data were collected is important. For example, training datasets collected in the morning or summer do not allow us to simulate buildings' envelope thermal information in the evening or winter, and vice versa. On the contrary, a physics-based approach is based on building materials and laws of thermodynamics. It can simulate building surface temperature at different times of a day and seasons by adjusting corresponding parameters.

In this study, we only evaluated the GAN model performance of simulating thermal images by implementing our approach on different datasets. We designed two evaluation metrics, MSE and SSIM values. The former is to evaluate the ability of simulating building envelope thermal information, and the latter is to evaluate the ability of simulating envelope appearances. As described in Section 5, we could reach some important conclusions: (1) Plotting all images' pairs of MSE and SSIM values shows a negative relationship between MSE and SSIM, namely one increasing while the other decreasing. If MSE and SSIM are investigated separately, we found out that a long tail and a Gaussian distribution can respectively describe MSE and SSIM values' distribution. (2) Using one model established by one building style to generate thermal images in another is not ideal. Both Table 1 and Figure 5 demonstrated that a case in which both training and testing datasets are in the same experiment (either city or campus experiment) performs better than other cases in which both datasets are in different experiments. It is wiser to use a training dataset that is similar to testing datasets for training the image translation models. (3) A GAN model built based on city datasets performs better than a model built based on campus datasets. This is because the city datasets have more complex buildings and features for the former model to learn. We suggest that researchers use datasets in which building information is richer and envelope structures are more complex as training datasets.

As described, the performance of deep learning using simulated images was not evaluated in this study. In future work, we plan to further evaluate the segmentation performance using simulated images by our method and current existing methods. When we compared our method with the current method, we did not use the same dataset since some researchers' methods were not open source. In the future, we will also consider integrating image generation with physics-based approaches to avoid their respective drawbacks.

Author Contributions: Conceptualization, Y.H. and L.S.; methodology, Y.H.; software, Y.H.; validation, Y.H.; resources, R.V.; data curation Y.H. and R.V., writing—original draft preparation, Y.H.; writing—review and editing, R.V. and L.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by China Scholarship Council (No.201606420072), Karlsruhe Institute of Technology (KIT) International Department, and the Deutscher Akademischer Austauschdienst, German Academic Exchange Service (DAAD)'s program.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors thank the CSC (China Scholarship Council), the KIT International Department, and the DAAD (Deutscher Akademischer Austauschdienst, German Academic Exchange Service) for their support. Specifically, the DAAD provided travel support. Furthermore, the authors thank Marinus Vogl, the remote pilot, and his drone company, Air Bavarian GmbH, for their support with equipment and service.

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

References

1. Chen, M.; Feng, A.; McAlinden, R.; Soibelman, L. Photogrammetric Point Cloud Segmentation and Object Information Extraction for Creating Virtual Environments and Simulations. *J. Manag. Eng.* **2020**, *36*, 04019046. [CrossRef]
2. Chen, M.; Feng, A.; Mccullough, K.; Prasad, B.; Mcalinden, R.; Soibelman, L. Semantic Segmentation and Data Fusion of Microsoft Bing 3D Cities and Small UAV-based Photogrammetric Data. *arXiv* **2020**, arXiv:2008.09648.
3. Omar, H.; Mahdjoubi, L.; Kheder, G. Towards an automated photogrammetry-based approach for monitoring and controlling construction site activities. *Comput. Ind.* **2018**, *98*, 172–182. [CrossRef]
4. He, K.; Gkioxari, G.; Dollar, P.; Girshick, R.B. Mask R-CNN. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 386–397. [CrossRef] [PubMed]
5. Wong, A.; Famuori, M.; Shafiee, M.J.; Li, F.; Chwyl, B.; Chung, J. YOLO Nano: A Highly Compact You Only Look Once Convolutional Neural Network for Object Detection. *arXiv* **2019**, arXiv:1910.01271.
6. Chen, L.-C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *40*, 834–848. [CrossRef]
7. Luo, C.; Sun, B.; Yang, K.; Lu, T.; Yeh, W.-C. Thermal infrared and visible sequences fusion tracking based on a hybrid tracking framework with adaptive weighting scheme. *Infrared Phys. Technol.* **2019**, *99*, 265–276. [CrossRef]
8. Liu, Z.; Zhang, W.; Zhao, P. A cross-modal adaptive gated fusion generative adversarial network for RGB-D salient object detection. *Neurocomputing* **2020**, *387*, 210–220. [CrossRef]
9. Zhai, S.; Shao, P.; Liang, X.; Wang, X. Fast RGB-T Tracking via Cross-Modal Correlation Filters. *Neurocomputing* **2019**, *334*, 172–181. [CrossRef]
10. Chen, H.; Li, Y.; Su, D. Multi-modal fusion network with multi-scale multi-path and cross-modal interactions for RGB-D salient object detection. *Pattern Recognit.* **2019**, *86*, 376–385. [CrossRef]
11. Lundervold, A.S.; Lundervold, A. An overview of deep learning in medical imaging focusing on MRI. *Z. Med. Phys.* **2019**, *29*, 102–127. [CrossRef] [PubMed]
12. Van Der Ploeg, T.; Austin, P.C.; Steyerberg, E.W. Modern modelling techniques are data hungry: A simulation study for predicting dichotomous endpoints. *BMC Med. Res. Methodol.* **2014**, *14*, 137. [CrossRef] [PubMed]
13. Shariq, M.H.; Hughes, B.R. Revolutionising building inspection techniques to meet large-scale energy demands: A review of the state-of-the-art. *Renew. Sustain. Energy Rev.* **2020**, *130*, 109979. [CrossRef]
14. Chen, M.; Feng, A.; Mcalinden, R.; Soibelman, L. Generating Synthetic Photogrammetric Data for Training Deep Learning based 3D Point Cloud Segmentation Models. *arXiv* **2020**, arXiv:2008.09647.
15. Li, C.; Xia, W.; Yan, Y.; Luo, B.; Tang, J. Segmenting Objects in Day and Night: Edge-Conditioned CNN for Thermal Image Semantic Segmentation. *IEEE Trans. Neural Networks Learn. Syst.* **2020**, *1*, 1–14. [CrossRef]
16. Han, J.; Chen, H.; Liu, N.; Yan, C.; Li, X. CNNs-Based RGB-D Saliency Detection via Cross-View Transfer and Multiview Fusion. *IEEE Trans. Cybern.* **2018**, *48*, 3171–3183. [CrossRef]
17. Chen, H.; Li, Y. Progressively Complementarity-Aware Fusion Network for RGB-D Salient Object Detection. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; Institute of Electrical and Electronics Engineers (IEEE): New York, NY, USA, 2018; pp. 3051–3060. [CrossRef]
18. Berg, A.; Ahlberg, J. Classification and temporal analysis of district heating leakages in thermal images. In Proceedings of the 14th International Symposium on District Heating and Cooling, Stockholm, Sweden, 7–9 September 2014.
19. Barreira, E.; Almeida, R.M.; Delgado, J.M.P.Q. Infrared thermography for assessing moisture related phenomena in building components. *Constr. Build. Mater.* **2016**, *110*, 251–269. [CrossRef]
20. Asdrubali, F.; Baldinelli, G.; Bianchi, F. A quantitative methodology to evaluate thermal bridges in buildings. *Appl. Energy* **2012**, *97*, 365–373. [CrossRef]
21. Fokaides, P.A.; Wongwises, S. Application of infrared thermography for the determination of the overall heat transfer coefficient (U-Value) in building envelopes. *Appl. Energy* **2011**, *88*, 4358–4365. [CrossRef]
22. Hou, Y.; Soibelman, L.; Volk, R.; Chen, M. Factors Affecting the Performance of 3D Thermal Mapping for Energy Audits in a District by Using Infrared Thermography (IRT) Mounted on Unmanned Aircraft Systems (UAS). In Proceedings of the 36th International Symposium on Automation and Robotics in Construction (ISARC) 2019, Banff, AB, Canada, 21–24 May 2019; pp. 266–273. [CrossRef]
23. Ilehag, R.; Schenk, A.; Huang, Y.; Hinz, S. KLUM: An Urban VNIR and SWIR Spectral Library Consisting of Building Materials. *Remote Sens.* **2019**, *11*, 2149. [CrossRef]
24. Bulatov, D.; Burkard, E.; Ilehag, R.; Kottler, B.; Helmholz, P. From multi-sensor aerial data to thermal and infrared simulation of semantic 3D models: Towards identification of urban heat islands. *Infrared Phys. Technol.* **2020**, *105*, 103233. [CrossRef]
25. Aguerre, J.P.; Nahon, R.; Garcia-Nevado, E.; La Borderie, C.; Fernández, E.; Beckers, B. A street in perspective: Thermography simulated by the finite element method. *Build. Environ.* **2019**, *148*, 225–239. [CrossRef]
26. Idczak, M.; Groleau, D.; Mestayer, P.G.; Rosant, J.-M.; Sini, J.-F. An application of the thermo-radiative model SOLENE for the evaluation of street canyon energy balance. *Build. Environ.* **2010**, *45*, 1262–1275. [CrossRef]
27. Roupioz, L.; Kastendeuch, P.; Nerry, F.; Colin, J.; Najjar, G.; Luhahe, R. Description and assessment of the building surface temperature modeling in LASER/F. *Energy Build.* **2018**, *173*, 91–102. [CrossRef]

28. Hénon, A.; Mestayer, P.G.; Lagouarde, J.-P.; Voogt, J. An urban neighborhood temperature and energy study from the CAPITOUL experiment with the SOLENE model. Part 1: Analysis of flux contributions. *Theor. Appl. Clim.* **2012**, *110*, 177–196. [CrossRef]
29. Kottler, B.; Burkard, E.; Bulatov, D.; Haraké, L. Physically-based Thermal Simulation of Large Scenes for Infrared Imaging. In VISIGRAPP 2019—Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, Prague, Czech Republic, 25–27 February 2019; VISIGRAPP: Prague, Czech Republic, 2019; Volume 1, pp. 53–64. [CrossRef]
30. Xiong, X.; Zhou, F.; Bai, X.; Xue, B.; Sun, C. Semi-automated infrared simulation on real urban scenes based on multi-view images. *Opt. Express* **2016**, *24*, 11345–11375. [CrossRef]
31. Hong, T.; Chen, Y.; Luo, X.; Luo, N.; Lee, S.H. Ten questions on urban building energy modeling. *Build. Environ.* **2020**, *168*, 106508. [CrossRef]
32. Allegrini, J.; Orehounig, K.; Mavromatidis, G.; Ruesch, F.; Dorer, V.; Evins, R. A review of modelling approaches and tools for the simulation of district-scale energy systems. *Renew. Sustain. Energy Rev.* **2015**, *52*, 1391–1404. [CrossRef]
33. Aguerre, J.P. Infrared Rendering for Thermography Simulation. 2020. Available online: https://github.com/jpaguerre/ThRend (accessed on 13 June 2020).
34. Aguerre, J.P.; Garcia-Nevado, E.; Miño, J.A.P.Y.; Fernández, E.; Beckers, B. Physically Based Simulation and Rendering of Urban Thermography. *Comput. Graph. Forum* **2020**, *39*, 377–391. [CrossRef]
35. And, B.B.; Garcia-Nevado, E. Urban Planning Enriched by Its Representations, from Perspective to Thermography. *Sustain. Vernac. Archit.* **2019**, 165–180. [CrossRef]
36. Hénon, A.; Mestayer, P.G.; Lagouarde, J.-P.; Voogt, J. An urban neighborhood temperature and energy study from the CAPITOUL experiment with the Solene model. Part 2: Influence of building surface heterogeneities. *Theor. Appl. Clim.* **2012**, *110*, 197–208. [CrossRef]
37. Xiong, B.; Elberink, S.O.; Vosselman, G. Building modeling from noisy photogrammetric point clouds. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2014**, *3*, 197–204. [CrossRef]
38. Luebke, D.; Reddy, M.; Cohen, J.D.; Varshney, A.; Watson, B.; Huebner, R. *Level of Detail for 3D Graphics*; Morgan Kaufmann: San Francisco, CA, USA, 2002; ISBN 9780080510118. Available online: https://www.elsevier.com/books/level-of-detail-for-3d-graphics/luebke/978-1-55860-838-2 (accessed on 1 December 2020).
39. Isola, P.; Zhu, J.-Y.; Zhou, T.; Efros, A.A. Image-to-Image Translation with Conditional Adversarial Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 5967–5976. [CrossRef]
40. Zhu, J.-Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 27–29 October 2017; Institute of Electrical and Electronics Engineers (IEEE): New York, NY, USA, 2017; pp. 2242–2251. [CrossRef]
41. Spencer, B.F.; Hoskere, V.; Narazaki, Y. Advances in Computer Vision-Based Civil Infrastructure Inspection and Monitoring. *Engineering* **2019**, *5*, 199–222. [CrossRef]
42. Jahanshahi, M.R.; Masri, S.F.; Padgett, C.W.; Sukhatme, G.S. An innovative methodology for detection and quantification of cracks through incorporation of depth perception. *Mach. Vis. Appl.* **2013**, *24*, 227–241. [CrossRef]
43. Liu, Y.-F.; Cho, S.; Spencer, B.F.; Fan, J.-S. Concrete Crack Assessment Using Digital Image Processing and 3D Scene Reconstruction. *J. Comput. Civ. Eng.* **2016**, *30*, 04014124. [CrossRef]
44. Paal, S.G.; Jeon, J.-S.; Brilakis, I.; Desroches, R. Automated Damage Index Estimation of Reinforced Concrete Columns for Post-Earthquake Evaluations. *J. Struct. Eng.* **2015**, *141*, 04014228. [CrossRef]
45. Yeum, C.M.; Dyke, S.J. Vision-Based Automated Crack Detection for Bridge Inspection. *Comput. Civ. Infrastruct. Eng.* **2015**, *30*, 759–770. [CrossRef]
46. Khaloo, A.; Lattanzi, D.; Cunningham, K.; Dell'Andrea, R.; Riley, M. Unmanned aerial vehicle inspection of the Placer River Trail Bridge through image-based 3D modelling. *Struct. Infrastruct. Eng.* **2018**, *14*, 124–136. [CrossRef]
47. Morgenthal, G.; Hallermann, N. Quality Assessment of Unmanned Aerial Vehicle (UAV) Based Visual Inspection of Structures. *Adv. Struct. Eng.* **2014**, *17*, 289–302. [CrossRef]
48. Tewkesbury, A.P.; Comber, A.J.; Tate, N.J.; Lamb, A.; Fisher, P.F. A critical synthesis of remotely sensed optical image change detection techniques. *Remote Sens. Environ.* **2015**, *160*, 1–14. [CrossRef]
49. Hussain, M.; Chen, D.; Cheng, A.; Wei, H.; Stanley, D. Change detection from remotely sensed images: From pixel-based to object-based approaches. *ISPRS J. Photogramm. Remote Sens.* **2013**, *80*, 91–106. [CrossRef]
50. Koch, C.; Paal, S.G.; Rashidi, A.; Zhu, Z.; König, M.; Brilakis, I. Achievements and Challenges in Machine Vision-Based Inspection of Large Concrete Structures. *Adv. Struct. Eng.* **2014**, *17*, 303–318. [CrossRef]
51. Xiong, X.; Adan, A.; Akinci, B.; Huber, D. Automatic creation of semantically rich 3D building models from laser scanner data. *Autom. Constr.* **2013**, *31*, 325–337. [CrossRef]
52. Armeni, I.; Sener, O.; Zamir, A.R.; Jiang, H.; Brilakis, I.; Fischer, M.; Savarese, S. 3D Semantic Parsing of Large-Scale Indoor Spaces. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; Institute of Electrical and Electronics Engineers (IEEE): New York, NY, USA, 2016; pp. 1534–1543. [CrossRef]
53. Golparvar-Fard, M.; Bohn, J.; Teizer, J.; Savarese, S.; Peña-Mora, F. Evaluation of image-based modeling and laser scanning accuracy for emerging automated performance monitoring techniques. *Autom. Constr.* **2011**, *20*, 1143–1155. [CrossRef]

54. Zhou, S.-J.; O'Neill, Z.; O'Neill, C. A review of leakage detection methods for district heating networks. *Appl. Therm. Eng.* **2018**, *137*, 567–574. [CrossRef]
55. Berg, A.; Ahlberg, J.; Berg, A. Classification of leakage detections acquired by airborne thermography of district heating networks. In Proceedings of the 2014 8th IAPR Workshop on Pattern Reconition in Remote Sensing, Stockholm, Sweden, 24 August 2014; Institute of Electrical and Electronics Engineers (IEEE): New York, NY, USA, 2014; pp. 1–4. [CrossRef]
56. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. In Proceedings of the Advances in Neural Information Processing Systems, Lake Tahoe, NV, USA, 3–6 December 2012; pp. 1097–1105. [CrossRef]
57. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
58. Zeiler, M.D.; Fergus, R. Visualizing and understanding convolutional networks, in European conference on computer vision. *arXiv* **2014**, arXiv:1311.2901, 818–833.
59. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014.
60. Iizuka, S.; Simo-Serra, E.; Ishikawa, H. Let there be color! *ACM Trans. Graph.* **2016**, *35*, 1–11. [CrossRef]
61. Larsson, G.; Maire, M.; Shakhnarovich, G. Learning Representations for Automatic Colorization. In *Lecture Notes in Computer Science*; Springer Science and Business Media LLC: Berlin/Heidelberg, Germany, 2016; Volume 9908, pp. 577–593. [CrossRef]
62. Zhang, R.; Isola, P.; Efros, A.A. Colorful Image Colorization. In *Computational Data and Social Networks*; Springer Science and Business Media LLC: Berlin/Heidelberg, Germany, 2016; Volume 9907, pp. 649–666. [CrossRef]
63. Mahdizadehaghdam, S.; Panahi, A.; Krim, H. Sparse Generative Adversarial Network. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), Seoul, Korea, 27–28 October 2019; pp. 3063–3071. [CrossRef]
64. Li, C.; Wand, M. Combining Markov Random Fields and Convolutional Neural Networks for Image Synthesis. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; Institute of Electrical and Electronics Engineers (IEEE): New York, NY, USA, 2016; pp. 2479–2486. [CrossRef]
65. Johnson, J.; Alahi, A.; Fei-Fei, L. Perceptual Losses for Real-Time Style Transfer and Super-Resolution. In *Computational Data and Social Networks*; Springer Science and Business Media LLC: Berlin/Heidelberg, Germany, 2016; Volume 9906, pp. 694–711. [CrossRef]
66. Mathieu, M.; Couprie, C.; LeCun, Y. Deep multi-scale video prediction beyond mean square error. In Proceedings of the 4th International Conference on Learning Representations, ICLR 2016—Conference Track Proceedings, San Juan, Puerto Rico, 2–4 May 2016; pp. 1–14.
67. Yoo, D.; Kim, N.; Park, S.; Paek, A.S.; Kweon, I.S. Pixel-Level Domain Transfer. In *Lecture Notes in Computer Science*; Springer Science and Business Media LLC: Berlin/Heidelberg, Germany, 2016; Volume 9912, pp. 517–532. [CrossRef]
68. Karacan, L.; Akata, Z.; Erdem, A.; Erdem, E. Learning to Generate Images of Outdoor Scenes from Attributes and Semantic Layouts. *arXiv* **2016**, arXiv:1612.00215.
69. Reed, S.; Akata, Z.; Mohan, S.; Tenka, S.; Schiele, B.; Lee, H. Learning what and where to draw, Advances in Neural Information Processing Systems. *arXiv* **2016**, arXiv:1610.02454, 217–225.
70. Li, C.; Wand, M. Precomputed Real-Time Texture Synthesis with Markovian Generative Adversarial Networks. In *Lecture Notes in Computer Science*; Springer Science and Business Media LLC: Berlin/Heidelberg, Germany, 2016; Volume 9907, pp. 702–716. [CrossRef]
71. Silva, E.A.; Panetta, K.; Agaian, S.S. Quantifying image similarity using measure of enhancement by entropy. In *Mobile Multimedia/Image Processing for Military and Security Applications 2007*; Defense and Security Symposium 2007: Orlando, FL, USA, 2007; Volume 6579, p. 65790U. [CrossRef]
72. AGandhi, S.; Kulkarni, C.V. MSE Vs SSIM. *Int. J. Sci. Eng. Res.* **2013**, *4*, 930–934. Available online: https://www.ijser.org/onlineResearchPaperViewer.aspx?MSE-Vs-SSIM.pdf (accessed on 15 December 2020).
73. Cordts, M.; Omran, M.; Ramos, S.; Rehfeld, T.; Enzweiler, M.; Benenson, R.; Franke, U.; Roth, S.; Schiele, B. The Cityscapes Dataset for Semantic Urban Scene Understanding. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; Institute of Electrical and Electronics Engineers (IEEE): New York, NY, USA, 2016; pp. 3213–3223. [CrossRef]