engineering proceedings

Proceeding Paper

# AI-Enhanced Detection of Thermal Anomalies in Urban Roofs via Drone-Assisted Infrared Thermography (UAV-IRT) †

**Xiaojia Zhang [1,*], Andrea Garzulino [1] and Elena Lucchi [2]**

[1] Department of Architecture and Urban Studies (DASTU), Politecnico di Milano, 20133 Milan, Italy; andrea.garzulino@polimi.it
[2] Department of Civil Engineering and Architecture, University of Pavia, 27100 Pavia, Italy; elena.lucchi@unipv.it
* Correspondence: xiaojia.zhang@polimi.it
† Presented at the 1st International Online Conference on Designs (Designs 2026), 9–10 February 2026; Available online: https://sciforum.net/event/Designs2026.

**Abstract**

This study investigates automated detection of rooftop thermal anomalies using Unmanned Aerial Vehicle-Based Infrared Thermography (UAV-IRT) and deep learning object detection models. UAV thermal images are used to identify potential thermal bridges on building rooftops. The experiment is framed as a preliminary workflow validation performed on a subset of the Karlsruhe reference dataset. The study evaluates four YOLO models, including YOLOv9, YOLOv10, YOLOv11, and YOLOv12. The results indicate that all models can detect major thermal anomaly regions in UAV thermal imagery. YOLOv9 achieves slightly higher detection accuracy in the tested dataset. These results demonstrate the feasibility of the proposed workflow under reference-dataset conditions, while validation on a dedicated historic-building dataset remains necessary before claiming applicability to heritage or historic urban environments.

**Keywords:** computer vision; object detection; infrared thermography; UAV; thermal anomaly detection; historic city

## 1. Introduction

Buildings account for about one-third of global energy consumption [1]. Improving building energy efficiency can reduce emissions and energy costs. The detection of thermal anomalies in building envelopes is important for improving energy performance. These anomalies include rooftop defects, thermal bridges, cracks, air leakage, and moisture infiltration that significantly affect the energy performance of buildings. Studies indicate that such defects can cause energy losses of about 20–34% [2]. The issue is more significant in historic urban environments, where dense building fabrics and aging rooftop structures often exhibit insufficient insulation, degraded materials, and construction techniques that predate modern energy standards [3]. These conditions often result in poor thermal performance but also highlight substantial potential for energy improvement through targeted retrofit strategies. Energy assessment in historic urban environments requires large-scale data collection and analysis of building envelopes, with rooftops representing one of the most critical components due to their exposure to climatic stress and their role in heat loss. Non-Destructive Testing (NDT) methods provide an effective approach because they allow rapid and non-invasive inspection of rooftop conditions. Infrared thermography (IRT) is widely used to detect rooftop thermal anomalies, such as insulation deficiencies, air leakage



Academic Editor: Surender Reddy Salkuti

Published: 20 July 2026

pathways, moisture infiltration, thermal bridges occurring at roof junctions, and structural discontinuities [4]. Recent developments have extended IRT through Unmanned Aerial Vehicles (UAVs), enabling efficient inspection of rooftops and improving access to complex geometries, steep pitches, and areas that are difficult to reach from ground-based surveys. UAV-based Infrared Thermography (UAV-IRT) has been increasingly combined with deep learning approaches for automated thermal image analysis [1,2,5]. Convolutional Neural Networks (CNNs) have shown strong performance in image recognition tasks. Among these models, the You Only Look Once (YOLO) architecture provides high detection accuracy and real-time processing capability, which makes it suitable for large-scale analysis of UAV-IRT [1]. Previous studies have primarily focused on rooftop thermal bridge detection in modern or standardized building envelopes. Mayer et al. conducted one of the earliest studies that applied deep learning to detect rooftop thermal bridges from aerial thermographic images [5], comparing several neural network architectures under both pretrained and non-pretrained settings. Their results showed that pretrained models achieved higher detection performance and supported the development of automated rooftop thermal bridge detection using UAV thermography. The authors also released the open dataset Thermal Bridges on Building Rooftops (TBBRv2), which contains 926 UAV images and 6927 annotated thermal bridges with RGB, thermal, and height information [6]. Waqas et al. investigated automated detection of thermal bridges in building envelopes using UAV-based multimodal aerial imagery [1]. Using the TBBRv2 dataset, they developed a YOLOv9-based model with additional CNN blocks to handle multimodal inputs, such as RGB, thermal, and LiDAR data. The model achieved a detection accuracy of 0.72, which increased to 0.763 after post-processing with an artificial neural network. Comparisons with Mask R-CNN and YOLOv7 indicated that the YOLO-based model achieved better detection performance on the TBBRv2 dataset. The trained model was also tested on another dataset to examine its generalization ability across different building conditions. These studies demonstrate the advantages of combining UAV-IRT with deep learning methods for building façade diagnostics at the urban scale. Despite these advances, thermal anomaly detection in historic rooftop at the urban level remains underexplored. Historic rooftops often contain heterogeneous construction materials, irregular geometries, and complex stratigraphy. These characteristics produce less regular thermal patterns than modern buildings. Existing UAV IRT and YOLO-based studies rarely address these conditions, which leaves a research gap and motivates future application-oriented validation in historic urban environments. In the present manuscript, however, the experimental validation is limited to a reference dataset from Karlsruhe. The historic-building dataset is treated as a future validation step rather than as evidence already tested in this study.

This study proposes an automated workflow for detecting rooftop thermal anomalies in historic urban environments. The workflow combines UAV-IRT with several YOLO models to assess their accuracy and reliability in the real-time detection of rooftop thermal anomalies from aerial thermal images. It also aims to provide a scalable framework for large-scale assessment. This framework further supports the integration of energy diagnostics into heritage conservation practice and helps connect building energy assessment with the preservation of historic urban environments.

## 2. Methods

This study uses a four-step workflow to preliminarily validate rooftop thermal anomaly detection using UAV-IRT (Figure 1). Step 1 focuses on data preparation. The dataset includes aerial thermal images collected from published sources and additional thermal images of historic buildings. The images are preprocessed, augmented, and labeled. The data are then formatted for YOLO-based training. Step 2 applies a YOLO-based CNN model to detect thermal anomalies. Step 3 evaluates model performance using standard metrics and further tests the trained model on additional datasets to assess its generalization capability. If the detection results are unstable or inaccurate, the network structure and modules are optimized. Step 4 applies the optimized model to automatically identify and visualize thermal anomalies in historic buildings at the urban scale.

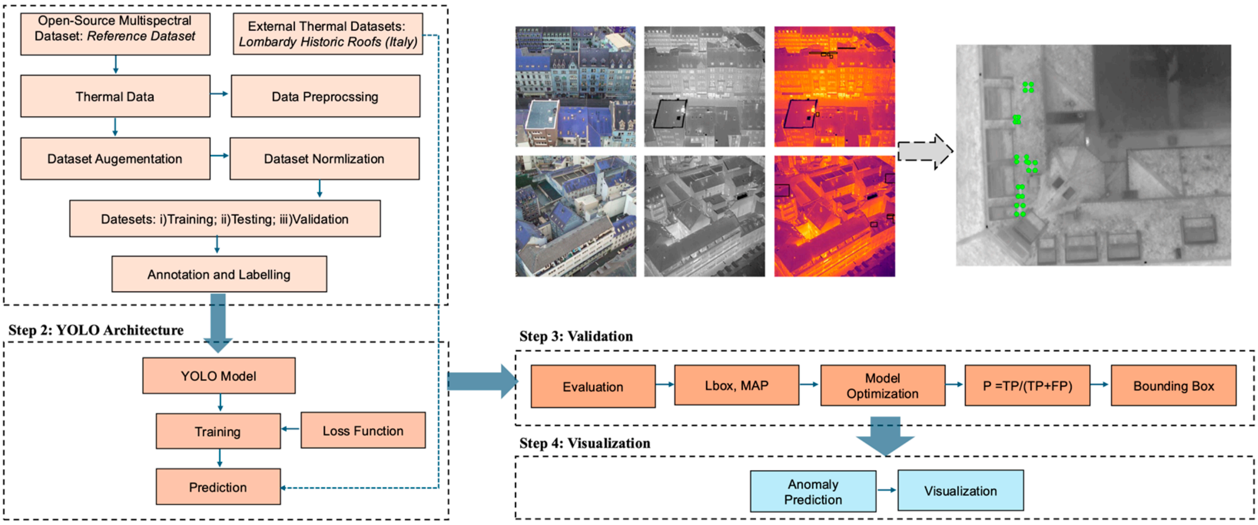


**Figure 1.** Proposed Workflow for automatically detecting thermal anomalies via UAV-IRT (source: Author's elaboration).

### 2.1. Data Generation

#### 2.1.1. Reference Dataset

This study first uses a publicly available aerial thermal dataset as a reference [6]. The thermal images were collected over rooftop areas in the city center of Karlsruhe, Germany. The surveyed area includes six urban blocks, with about twenty buildings in each. These buildings differ in age, form, and construction type. This variation provides different rooftop geometries and thermal patterns that are suitable for studying rooftop heat loss. The data were collected on 19 March 2019, between 07:00 and 08:00 (UTC + 02:00). The early morning time was chosen to limit the effect of solar radiation on surface temperature. During the flight, the air temperature ranged from 3.78 °C to 4.97 °C, and the relative humidity ranged from 80% to 98%. No rain occurred during the flight, but about 2.3 mm of rainfall had been recorded within the previous 48 h. These weather conditions made it possible to take stable thermal pictures of building rooftops. The images were captured using a DJI M600 UAV (DJI, Shenzhen, China) equipped with a Zenmuse XT2 camera (DJI, Shenzhen, China) and a FLIR Tau 2 thermal sensor (Teledyne FLIR, Wilsonville, OR, USA). The full dataset contains about 926 thermal images. This study uses a subset of 322 images for initial model training and testing. The selected images show rooftops from different viewing angles and flight heights. Many rooftops appear in several images because the UAV flight created a high level of image overlap. Some buildings appear in up to twenty images from different perspectives. The original thermal images have a resolution of 640 $\times$ 512 pixels. All images were resized to 3000 $\times$ 4000 pixels and then cropped to remove blank borders. The final images have a size of 2400 $\times$ 3400 pixels. No dedicated historic-building dataset is used for the experimental results reported in this paper. Future work will construct and annotate a historic-building dataset to validate the transferability of the proposed workflow to heritage contexts, where building materials, roof structures, and thermal characteristics often differ from those of modern buildings.

#### 2.1.2. Labelling

The annotation process used LabelImg version 1.8.6, which allows manual drawings of bounding boxes around regions of interest (ROIs) [7]. In this study, the ROIs correspond to thermal anomalies in the thermal images. The manual annotation produced structured ground truth labels. These labels are required for training and evaluating the YOLO-based detection model. All annotations were manually performed by the first author and reviewed by the co-authors with expertise in IRT and building diagnostics. Uncertain cases were resolved through discussion until consensus was reached, while images with ambiguous thermal patterns or indistinct boundaries were excluded. During the annotation process, the annotator identified regions exhibiting clear temperature anomalies. Each bounding box was positioned closely around the visible thermal anomaly, while areas showing weak or ambiguous temperature variation were excluded. The objective was to capture the core thermal pattern rather than the surrounding background. Some images contained partial occlusions. In these cases, foreground architectural elements obscured part of the thermal anomaly. When the obstruction was limited, and the anomaly remained interpretable, the annotation still encompassed the entire thermal bridge area. Thermal bridges associated with different building components were labeled as separate instances. For example, a poorly insulated window reveal could appear within a larger roof area affected by heat loss. In this situation, the window reveal and the surrounding roof anomaly were annotated independently. Bounding boxes were defined without overlap so that each annotated region represented a single thermal anomaly. Some minor anomalies were excluded when their boundaries lacked sufficient precision or when the image did not reliably reveal the thermal pattern.

#### 2.1.3. Dataset Generation

The annotation process produced a total of 2882 labeled instances. Each instance corresponds to a thermal bridge or another localized temperature anomaly visible in the infrared imagery. The annotated images form the effective dataset used for model development because supervised object detection requires labeled samples during training. The annotated dataset was divided into a training set and a validation set using an 8:2 ratio. The training set contains 258 images and 2293 annotations. The validation set includes 64 images and 589 annotations. The training set provides the samples used for model training, while the validation set is reserved for monitoring model performance during the training process. Although the 8:2 split is commonly adopted, the train–validation partition represents an experimental setting rather than a universally optimal standard. Previous studies have shown that different data partitioning strategies may influence model performance and evaluation outcomes [8,9]. Therefore, the reported results should be interpreted with this consideration in mind, and future studies could further assess model robustness using repeated random splits or cross-validation techniques.

The annotation of thermal anomalies is a time-intensive process because the boundaries of these regions are often indistinct and require careful interpretation of infrared imagery. At this stage, a subset of the reference images was selected to provide an initial validation of the proposed workflow rather than a definitive domain validation for historic buildings. The reference images and the planned historic-building images will be annotated in future work to expand the training dataset and enable a more comprehensive evaluation of model performance and transferability.

### *2.2. YOLO Architecture*

The YOLO series has developed quickly in recent years. Several new versions have been proposed to improve detection accuracy and real-time performance. YOLOv9 focuses

https://doi.org/10.3390/engproc2026138015

on a common problem in deep neural networks, which is the loss of useful information during training. The model keeps important feature information while the model learns from the data [10]. YOLOv10 then improves the network structure used for feature extraction. It uses an upgraded CSPNet backbone and improved PANet layers to support multiscale feature fusion. The model also adopts a dual-head design. During training, the one-to-many head generates several predictions for each object and provides stronger supervision signals. During inference, the one-to-one head produces a single prediction for each object, so the model can avoid the use of non-maximum suppression [11]. YOLOv11 continues this development and focuses on stronger feature representation. The architecture introduces modules (e.g., Cross Stage Partial with Kernel size 2, C3K2), which help the model capture information at different spatial scales and help the network focus on important regions in the image [12]. These improvements are useful when objects appear in complex scenes or when they are of different sizes. YOLOv12 is the most recent version and introduces several structural changes to improve both accuracy and efficiency. The model keeps the backbone–neck–head design used in earlier networks, as illustrated in Figure 2. At the same time, the architecture adds lightweight attention mechanisms to improve feature learning. In the backbone network, modules such as C3K2 and Area Attention-enhanced Cross-Stage Fusion (A2C2f) support feature extraction and expand the network's receptive field. The model also includes an area attention module. This module allows the network to capture spatial relationships in the image while keeping the computational cost low [13]. In recent computer vision studies, attention mechanisms and efficient feature aggregation have become important directions for improving detection models. These design ideas also appear in YOLOv12, which shows how the model combines efficiency with stronger feature representation for real-time detection tasks.

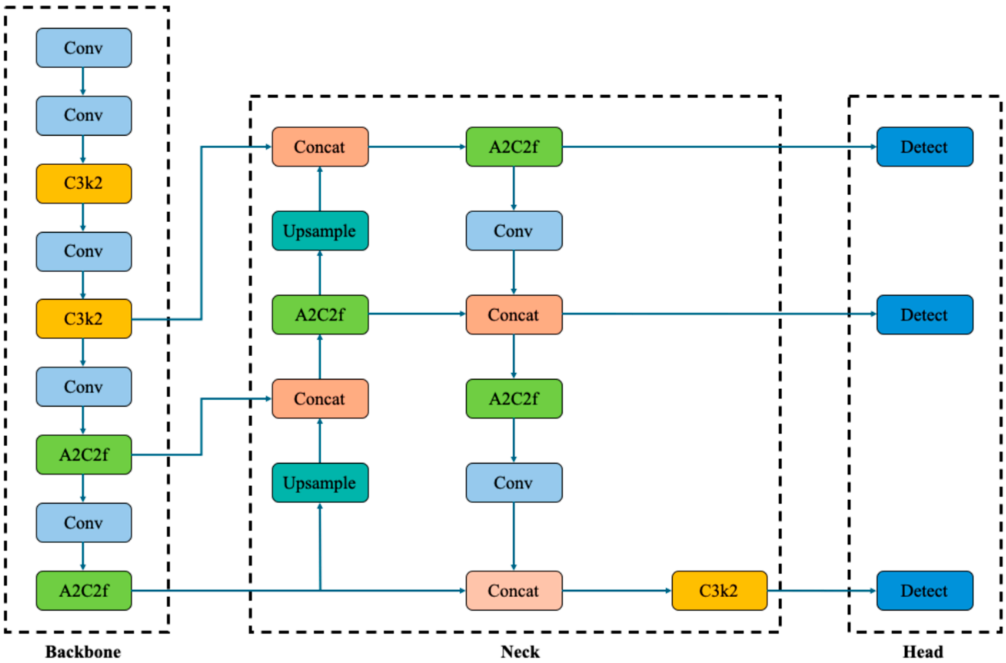


**Figure 2.** YOLO v12 Architecture (source: Author's elaboration).

### 2.3. Validation

Object detection performance in this study is described using four outcomes: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). These outcomes compare the model predictions with the ground truth labels. TP indicates that a thermal anomaly exists, and the model correctly detects it. TN indicates that no anomaly exists and the model correctly predicts its absence. FP refers to cases where the model predicts an anomaly that is not present. FN refers to cases where an existing anomaly is not detected. Precision, Recall, Intersection over Union (IoU), and mean Average Precision at an IoU threshold of 0.5 (mAP0.5) are used to evaluate model performance. The formulas for these metrics are presented in Equations (1)–(4). IoU measures the overlap between the predicted bounding box and the ground truth bounding box. IoU is calculated as the ratio between the intersection area and the union area of the predicted bounding box $A_p$ and the ground truth bounding box $A_{gt}$. The mAP0.5 metric is calculated from the precision–recall curve at an IoU threshold of 0.5.

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

$$IoU = \frac{Area\ of\ Overlap}{Area\ of\ Union} = \frac{A_p \cap A_{gt}}{A_p \cup A_{gt}} \quad (3)$$

$$\text{mAP0.5} = \frac{1}{N}\sum_{i=1}^{N} \text{AP}_{i,\text{IoU}=0.5} \quad (4)$$

## 3. Results and Discussion

Four YOLO models, including YOLOv9-c, YOLOv10-n, YOLOv11-n, and YOLOv12-n, are used to validate the proposed workflow. Their detection performance is evaluated and compared. Model training requires a clear definition of key hyperparameters. In this study, the model parameter specifies the network structure or the initial weights. The model can be loaded from a pre-trained weight file (.pt) or from a configuration file (.yaml). The dataset configuration file defines the paths and category information for the training and validation data so that the model can correctly load the dataset. The training process runs for 200 epochs. The patience parameter is set to 100 to stop training when the validation performance no longer improves. The batch size is set to 32 so that each training step uses enough samples while keeping GPU memory usage stable. The input image size is fixed at 640 × 640 pixels to match the YOLO input format and maintain a balance between detection accuracy and computational cost. Model training is conducted on a workstation equipped with an NVIDIA RTX 6000 Ada GPU with 48 GB VRAM. All models start from pre-trained weights to improve training efficiency. The optimizer is selected automatically by the framework. The initial learning rate is set to 0.01, and the learning rate gradually decreases during training according to the learning rate schedule.

Table 1 presents the detection performance of four YOLO models. YOLOv9 shows the highest detection accuracy in this experiment. The model achieves a precision of 0.905 and a recall of 0.854. The model also obtains the highest mAP@0.5 value of 0.893 and the highest mAP@0.5–0.95 value of 0.540. These results indicate that YOLOv9 is better at finding thermal anomalies on rooftops via UAV-IRT than the other models in this dataset. The higher recall also indicates that YOLOv9 produces fewer missed detections. This behavior is important in thermal anomaly inspection, where missed anomalies may correspond to potential thermal bridges on building envelopes. YOLOv10 and YOLOv11 present slightly lower detection accuracy but faster inference speed. Both models complete inference in about 4.4 milliseconds (ms) per image. This speed is faster than YOLOv9, which requires 6.9 ms per image. The difference suggests that the newer models emphasize computational efficiency. Such efficiency may be useful in real-time inspection scenarios, for example, when UAV systems collect large numbers of thermal images during urban surveys. YOLOv12 shows balanced performance between detection accuracy and inference speed. The model reaches a precision of 0.875 and a recall of 0.844. The mAP@0.5 value is 0.879, and the mAP@0.5–0.95 value is 0.523. The inference time is 5.0 ms per image. These results indicate that YOLOv12 maintains stable detection performance while keeping computational cost relatively low. The comparison also reveals an intriguing observation. The most recent model does not always achieve the highest detection accuracy in this experiment. YOLOv9 performs slightly better than YOLOv12 in several metrics. A possible explanation is that the dataset includes a relatively limited number of training images and comparatively simple thermal patterns. Many rooftops within the dataset exhibit similar material compositions and structural configurations. Under these conditions, larger or more complex model architectures may not yield substantial performance gains. Accordingly, these results should be interpreted as a comparison of model behavior on the Karlsruhe reference subset, not as evidence of direct operational performance in historic-building contexts. Historic buildings typically display greater variability in roof structures, construction materials, and thermal behavior than contemporary buildings. Future research will expand the dataset by incorporating dedicated examples from historic buildings to investigate how different model architectures respond to more complex thermal patterns and conservation-related building conditions.

**Table 1.** Model performance comparison (source: Author's elaboration).

| Model | Precision | Recall | mAP@0.5 | mAP@0.5–0.95 | Inference Time |
|---|---|---|---|---|---|
| YOLO v9 | 0.905 | 0.854 | 0.893 | 0.540 | 6.9 ms |
| YOLO v10 | 0.862 | 0.830 | 0.886 | 0.513 | 4.4 ms |
| YOLO v11 | 0.867 | 0.835 | 0.882 | 0.514 | 4.4 ms |
| YOLO v12 | 0.875 | 0.844 | 0.879 | 0.523 | 5.0 ms |

Figure 3 presents a visual comparison of rooftop thermal anomaly detection results generated by four YOLO models. The first row shows the input thermal images captured by the UAV-IRT. The subsequent rows show the detection outputs from YOLOv9, YOLOv10, YOLOv11, and YOLOv12. The green bounding boxes in the input images represent the ground-truth annotations of the thermal anomalies. Blue bounding boxes indicate the detected thermal anomaly regions, and the numbers represent the confidence scores of the predictions. All four models detect the main thermal anomaly areas visible in the thermal images. These areas appear as localized high-temperature regions on the rooftops. The models successfully identify several large thermal bridge zones located along roof ridges, roof edges, and structural joints. These locations appear consistently across the detection results of all models. However, all four models exhibit varying degrees of missed detections when identifying smaller thermal anomaly regions. Compared with the input thermal images, thermal anomalies distributed around densely arranged rooftop windows (left column) are only partially detected by all four models, indicating that closely spaced thermal anomalies remain challenging for automatic object detection. As the scene complexity increases (middle and right columns), the differences among the models become more apparent. Rooftops containing multiple closely spaced thermal anomalies, complex structural layouts, and thermal patterns located near the image boundaries further increase the detection difficulty. Consequently, all four models exhibit occasional missed detections in these challenging regions, whereas no obvious false-positive detections are observed.

YOLOv9 generally produces bounding boxes that more closely align with the ground-truth annotations. YOLOv10 and YOLOv11 show similar detection patterns but miss several smaller anomalies in some scenes, particularly those located near the image boundaries. YOLOv12 exhibits detection patterns that are broadly consistent with those of the other models, although slight variations in the number and spatial distribution of detected thermal anomalies can be observed across different scenes. These visual results are consistent with the quantitative evaluation results reported in Table 1. Although YOLOv9 demonstrates slightly better overall detection performance than the other models, the qualitative detection capability of all four YOLO models remains broadly comparable. The observed differences are primarily associated with the detection of small-scale thermal anomalies rather than the major thermal bridge regions. These missed detections are primarily associated with the limited spatial extent of the anomalies, weak thermal contrast with the surrounding roof surfaces, and the presence of densely distributed rooftop structures, which reduce the discriminative thermal features available for reliable object detection. These findings indicate that the detection performance of thermal anomaly analysis depends not only on the model architecture but also on the characteristics of the thermal imagery and the structural features of the rooftops.

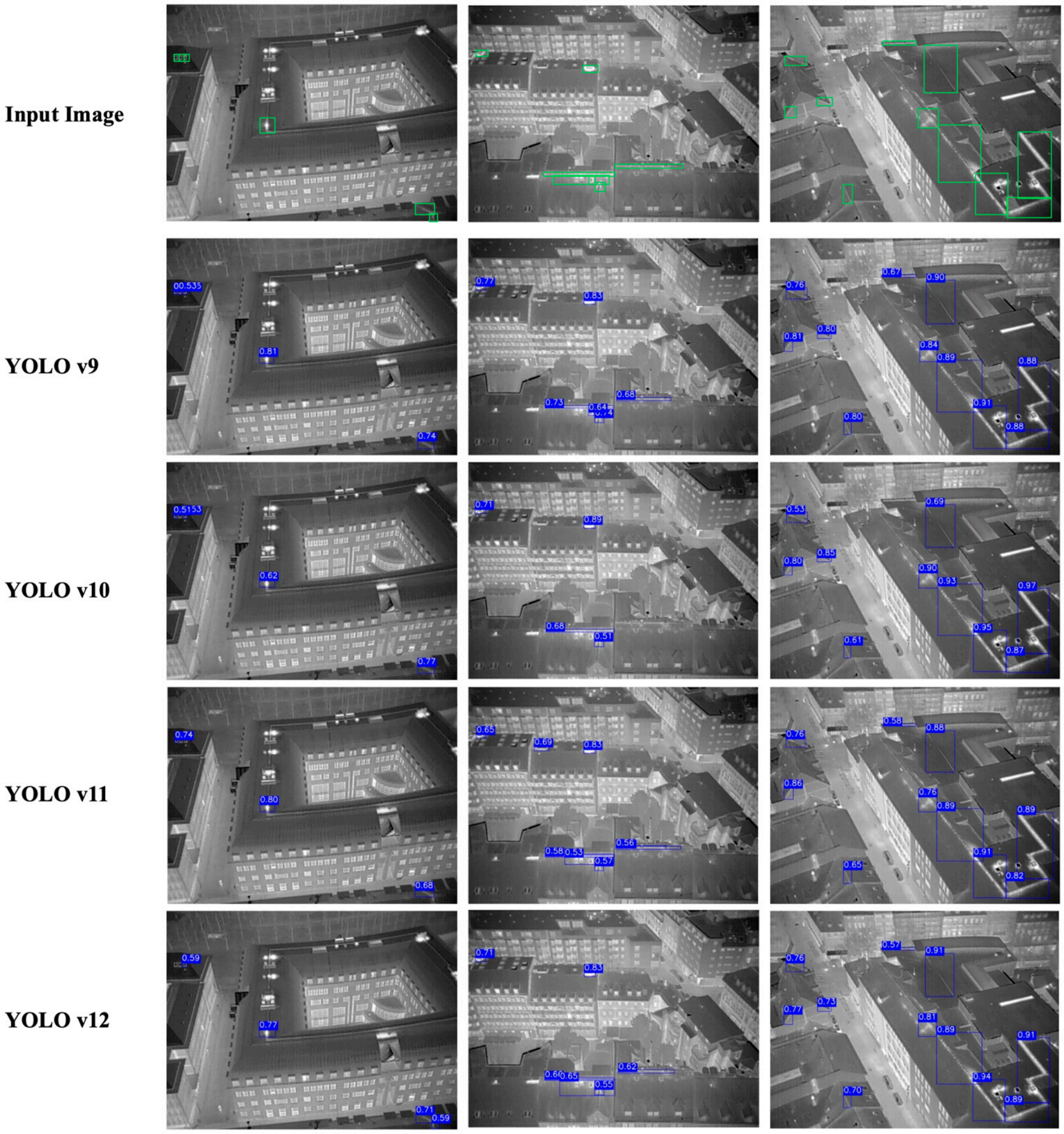


**Figure 3.** Comparison of rooftop thermal anomaly detection results produced by four YOLO models. Green bounding boxes in the input images indicate the ground-truth annotations of thermal anomalies. Blue bounding boxes indicate the detected thermal anomalies, and the numerical labels indicate the corresponding confidence scores (source: Author's elaboration).

## 4. Conclusions

This study presents a preliminary validation of a computer vision workflow for detecting rooftop thermal anomalies from UAV infrared thermal imagery. The workflow uses YOLO-based object detection models to identify thermal bridges in aerial thermal images. This study tests four models: YOLOv9, YOLOv10, YOLOv11, and YOLOv12. The experimental results indicate that all models can detect thermal anomaly regions on building rooftops. The quantitative results indicate that YOLOv9 achieves slightly higher detection accuracy in the tested dataset, while the other models provide comparable detection performance with faster inference speeds. All four YOLO models reliably identify the major thermal bridge regions, whereas their primary differences are observed in the detection of small-scale thermal anomalies. These missed detections are mainly attributed to small anomaly size, weak thermal contrast, and complex rooftop structures, highlighting the influence of thermal characteristics on detection performance. The results show that deep learning-based object detection methods can support automated analysis of rooftop thermal conditions in UAV-IRT. This approach provides a practical tool for large-scale inspection of building envelopes in urban environments. However, the present study is subject to several limitations. The current training dataset consists only of a subset of the available thermal image archive. The limited number of annotated samples constrains the diversity of rooftop conditions represented during the training process. Many rooftops in the dataset share comparable construction materials and structural configurations, which may reduce the variability of thermal patterns captured by the model. This situation may affect the capacity of the model to learn more complex thermal behaviors that occur under different building conditions. In addition, thermal anomalies in infrared imagery often appear as diffuse temperature gradients rather than clearly defined geometric shapes. This characteristic increases the difficulty of precise localization for object detection models that rely on rectangular bounding boxes. Future work will expand the current dataset by completing the annotation of the full thermal image archive. The expanded dataset will support additional training and validation to improve model robustness and reduce uncertainty in detection results. Future work will also further assess the robustness of the proposed workflow using different train–validation splits and cross-validation techniques. Future research will also construct and annotate a dedicated UAV infrared thermal dataset that focuses on rooftop thermal anomalies in historic buildings. Historic buildings often contain more complex roof structures, diverse construction materials, and heterogeneous thermal behavior. A dataset that represents these characteristics will allow the heritage-oriented workflow to be tested systematically and will support further research on the generalization ability of computer vision methods for thermal diagnostics of historic building envelopes.

**Author Contributions:** Conceptualization, X.Z., A.G. and E.L.; methodology, X.Z.; software, X.Z. and A.G.; data curation, X.Z. and A.G.; writing—original draft preparation, X.Z.; writing—review and editing, A.G. and E.L.; supervision, A.G. and E.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The datasets presented in this article are not readily available due to technical and time limitations.

**Acknowledgments:** Xiaojia Zhang would like to thank the financial support from the China Scholarship Council (CSC) under the grant CSC No. 202309210009.

**Conflicts of Interest:** The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## Abbreviations

The following abbreviations are used in this manuscript:

| | | | |
|---|---|---|---|
| A2C2f | Area Attention-enhanced Cross-Stage Fusion | QIRT | Quantitative Infrared Thermography |
| CNN | Convolutional Neural Network | ROIs | Regions of Interest |
| C3K2 | Cross Stage Partial with kernel size 2 | UAV | Unmanned Aerial Vehicle |
| IRT | Infrared Thermography | UAV-IRT | Drone-assisted Infrared Thermography |
| NDT | Non-Destructive Testing | YOLO | You Only Look Once |

## References

1. Waqas, A.; Araji, M.T.; Sherif, S.S. A Novel Method for Monitoring Material Behaviour in Urban Buildings Using Aerial Multi-Modality Imaging. *Build. Environ.* **2025**, *282*, 113300. [CrossRef]
2. Waqas, A.; Araji, M.T. Machine Learning-Aided Thermography for Autonomous Heat Loss Detection in Buildings. *Energy Convers. Manag.* **2024**, *304*, 118243. [CrossRef]
3. Zhang, X.; Garzulino, A.; Lucchi, E. Optimized Workflow for Drone-Assisted Quantitative Thermography in Urban Scale Energy Management. In *Proceedings of the Journal of Physics: Conference Series*; IOP Publishing: Bristol, UK, 2025; Volume 3140, p. 062015.
4. Lucchi, E. Applications of the Infrared Thermography in the Energy Audit of Buildings: A Review. *Renew. Sustain. Energy Rev.* **2018**, *82*, 3077–3090. [CrossRef]
5. Mayer, Z.; Kahn, J.; Hou, Y.; Götz, M.; Volk, R.; Schultmann, F. Deep Learning Approaches to Building Rooftop Thermal Bridge Detection from Aerial Images. *Autom. Constr.* **2023**, *146*, 104690. [CrossRef]
6. Mayer, Z.; Kahn, J.; Hou, Y.; Beiersdörfer, T.; Blumenröhr, N.; Götz, M.; Volk, R. *Thermal Bridges on Building Rooftops-Hyperspectral (RGB + Thermal + Height) Drone Images of Karlsruhe, Germany, with Thermal Bridge Annotations*; Zenodo: Geneva, Switzerland, 2022.
7. Ilani, M.A.; Amini, L.; Karimi, H.; Kuhshuri, M.S. CNN-Based Labelled Crack Detection for Image Annotation. *arXiv* **2024**, arXiv:2408.11250.
8. Wang, S.; Hae, H.; Kim, J. Development of Easily Accessible Electricity Consumption Model Using Open Data and GA-SVR. *Energies* **2018**, *11*, 373. [CrossRef]
9. Wang, S. Class-Aware Temporal and Contextual Contrastive Framework for Semi-Supervised Automated Fault Detection and Diagnosis in Air Handling Units. *Energy Build.* **2026**, *358*, 117233. [CrossRef]
10. Wang, C.-Y.; Yeh, I.-H.; Mark Liao, H.-Y. Yolov9: Learning What You Want to Learn Using Programmable Gradient Information. In *Proceedings of the European Conference on Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2024; pp. 1–21.
11. Wang, A.; Chen, H.; Liu, L.; Chen, K.; Lin, Z.; Han, J.; Ding, G. Yolov10: Real-Time End-to-End Object Detection. *Adv. Neural Inf. Process. Syst.* **2024**, *37*, 107984–108011. [CrossRef]
12. Khanam, R.; Hussain, M. Yolov11: An Overview of the Key Architectural Enhancements. *arXiv* **2024**, arXiv:2410.17725.
13. Tian, Y.; Ye, Q.; Doermann, D. Yolov12: Attention-Centric Real-Time Object Detectors. *arXiv* **2025**, arXiv:2502.12524.