

ORIGINAL RESEARCH ARTICLE

Examining the nuances of Huizhou architecture and building decoration elements within the framework of rural development and urban aesthetics through the application of object detection and explicative analysis

Linyu Liu^{1,2,*}, Raziah Ahmad³, Suriati Ahmad⁴, Xuejie Jiang¹

¹ Department of Built Environment Studies and Technology, College of Built Environment, University Technology MARA, Kampus Seri Iskandar, Seri Iskandar 32610, Perak Darul Ridzuan, Malaysia

² Anhui University of Technology, Ma'anshan 2430323, China

³ Centre of Studies for Landscape Architecture Faculty of Architecture, Planning and Surveying Built Environment Complex Universiti Teknologi MARA Puncak Alam Campus, Bandar Puncak Alam 42300, Selangor, Malaysia

⁴ Landscape Architecture Program Department of Built Environment Studies and Technology Faculty of Architecture, Building and Surveying, UiTM Perak Branch, Seri Iskandar 32610, Perak Darul Ridzuan, Malaysia

* Corresponding author: Linyu Liu, lly2001182021@163.com

ABSTRACT

This scholarly investigation immerses itself in the intricate domain of Huizhou architecture and building decoration elements, which represent a distinct traditional Chinese architectural style. The study explores the evolution of Huizhou-style architecture and its ornamental techniques while delving into the visual aspects of contemporary urban architecture influenced by this unique style. Employing advanced technologies such as computer vision and explainable artificial intelligence (AI), the research aims to contribute to the preservation and documentation of China's cultural heritage, specifically focusing on the intricate designs and distinctive building decorations inherent in Huizhou-style architecture. To methodically curate a diverse image dataset, a combination of automated annotation tools and manual labeling was utilized, establishing a robust benchmark for the exploration and comprehension of Huizhou architectural elements. The YOLOv7 model underwent retraining on this dataset, exhibiting noteworthy enhancements in precision, recall, and Mean Average Precision (mAP), surpassing the performance of the pre-trained model. In addition, the study introduces SHAPE, an explainable AI tool designed for interpretability, providing detailed insights into the decision-making process. This not only bolsters the reliability of our results but also enriches our comprehension of Huizhou-style architecture. This multidimensional approach not only propels the field of computer vision but also makes significant strides in the preservation of China's cultural legacy. By integrating cutting-edge technologies with a meticulous exploration of architectural elements, this research fosters a deeper understanding of Huizhou architecture and its role in shaping the visual landscape of both rural and urban environments.

Keywords: Huizhou style architecture; building decoration; rural development; YOLO; explainable AI; object detection

1. Introduction

The architectural tapestry of Huizhou style design stands as a vibrant tribute to China's cultural legacy, boasting distinctive features deeply ingrained in tradition. This study intricately explores the nuances of Huizhou-style architecture and building decoration methods within the context of rural development and urban aesthetics. Leveraging the tools of object detection and explicative analysis, this research adopts a contemporary perspective to unravel the essence of these traditional elements, bridging the historical and modern realms. The paper aims to illuminate the evolving architectural practices, emphasizing the symbiotic relationship between tradition and technological innovation^[1,2].

ARTICLE INFO

Received: 5 February 2024
Accepted: 21 February 2024
Available online: 27 May 2024

COPYRIGHT

Copyright © 2024 by author(s).
Journal of Autonomous Intelligence is
published by Frontier Scientific Publishing.
This work is licensed under the Creative
Commons Attribution-NonCommercial 4.0
International License (CC BY-NC 4.0).
<https://creativecommons.org/licenses/by-nc/4.0/>

Huizhou-style architectural heritage has garnered increasing interest in both academic and industrial spheres. Previous research has predominantly concentrated on the historical and cultural significance of Huizhou architecture, emphasizing its impact on the built environment in both rural and urban landscapes. Scholars have scrutinized the ornamental techniques intrinsic to traditional Chinese architecture, underscoring the aesthetic and cultural value embedded in these designs. Additionally, studies have explored the challenges of integrating modern technologies into traditional architectural practices, seeking a harmonious coexistence of cultural heritage and contemporary advancements^[3-5].

Numerous investigations have been conducted on the integration of computer vision and AI within various domains, including architecture. Techniques such as object detection, image analysis, and spatial mapping have been employed to comprehend and replicate traditional architectural styles. Furthermore, explainable AI has gained prominence in architecture, particularly in decision-making processes, aiming for transparency and interpretability in design rooted in cultural principles^[6-9].

This paper contributes to the existing body of knowledge by adopting an interdisciplinary approach that converges the realms of Huizhou-style architecture, building decoration, rural development, urban visuals, computer vision, and explainable AI. The study endeavors to offer a comprehensive understanding of the intricate interplay between tradition and technology within the architectural domain, setting the stage for sustainable, culturally rich, and technologically informed architectural practices.

The primary contributions of this research paper can be summarized as follows:

- a) Comprehensive image dataset: The study contributes to the creation of a diverse image dataset specifically focused on Huizhou style architecture. Meticulously curated through a combination of automated annotation and manual labeling, this dataset serves as a valuable resource for exploring and understanding the intricacies of Huizhou architectural elements.
- b) Enhanced model performance: The retraining of the YOLOv7 model on the curated dataset results in a significant improvement in performance metrics, including precision, recall, and Mean Average Precision (mAP). The enhanced model showcases its capability to accurately identify and classify relevant objects within the context of Huizhou style architecture.
- c) Explainable AI insights: The integration of SHAPE as an explainable AI tool contributes to the interpretability of the model's decision-making process. SHAPE provides detailed insights into why certain architectural features are prioritized in the model's predictions, enhancing the transparency and trustworthiness of the

results.

The paper is structured as follows: In section 2, we delve into the literature survey, providing an overview of the existing research in the field. Section 3 outlines the proposed system, detailing its design and implementation. Section 4 focuses on the results and facilitates a comprehensive discussion. Lastly, in section 5, we conclude from our research findings and outline potential directions for future research.

Related work

Starzyńska-Grześ et al.^[10] conduct an in-depth exploration into the analysis of buildings and the built environment using computer vision. This study aims to evaluate the relevance of such research in the context of architectural studies, specifically focusing on Huizhou-style architecture and building decoration elements. The transition from an architecture-centric to a computer science-centric approach is scrutinized, shedding light on the associated research challenges, objectives, methodologies, and practical applications. Notably, the authors address challenges encountered in applying computer vision to architectural analysis, including the scarcity of large datasets of architectural images, the diversity of architectural styles, and the necessity for domain knowledge to interpret the outcomes of computer vision algorithms^[10].

In a related vein, Xia et al.^[11] propose a machine learning algorithm for predicting and classifying styles in residential buildings, emphasizing the challenges posed by the intricate and ambiguous nature of architectural styles. This paper responds to the complexities by introducing a systematic method for classifying and forecasting housing building styles^[11].

Ren^[12] contribute to the discourse on rural development, particularly in China, by challenging the prevalent top-down approach and advocating for a more participatory and bottom-up strategy. This approach involves the active engagement of residents in the design and implementation of architectural projects, aligning with the theme of examining Huizhou architecture within the framework of rural development. Drawing on theoretical and empirical sources, the study highlights case studies of architectural projects in Chinese rural villages that embrace hybrid approaches^[12].

Similarly, Xu et al.^[13] delve into the realm of indoor environmental art design, developing a hybrid conformal prediction algorithm framework (HCPAF). This framework, with its roots in Huizhou-style architecture, combines conformal prediction and hybrid prediction to enhance accuracy and robustness in design. The study demonstrates HCPAF's capability, achieving a remarkable accuracy of 98% and an execution rate of 92%^[13].

Mirra^[14] explore the utilization of artificial intelligence (AI) in generating design ideas for indoor environmental art, acknowledging both the creative potential and the irreplaceable role of human designers in the process^[14].

As et al.^[15] highlight the transformative impact of AI on various industries, including urban planning and design. This aligns with the theme of urban aesthetics and the exploration of AI applications in the context of Huizhou-style architecture. The study emphasizes the potential of AI to address challenges in urban planning, such as traffic congestion, air pollution, and social inequality^[15].

Chen and Xiang^[16] delve into the intricate relationship between tourism, commodification, and rural transformation, providing insights into the impact of tourism-led commodification on rural communities. This perspective complements the examination of Huizhou architecture within the broader framework of rural development and urban aesthetics^[16].

Wu et al.^[17] and Xiao et al.^[18] explore computational techniques to analyze and understand urban environments, demonstrating the potential of these techniques in providing new insights. These studies align with the theme of object detection and explicative analysis, showcasing applications in visualizing thermal

environments and marketing tourism destinations^[17-19].

Huang^[20] contributes to the literature on sustainable tourism development by presenting the design of Chinese rural homestay inns based on the concept of symbiosis. This aligns with the examination of Huizhou architecture and its integration with sustainable practices in rural settings^[20].

Maskeliunas et al.^[21] make a valuable contribution to the intersection of computer vision and architectural heritage documentation by building a façade style classification from UAV imagery using a deep learning algorithm. The study showcases the effectiveness of their approach on a dataset of over 8000 UAV images of buildings, providing insights relevant to the examination of Huizhou architecture and building decoration elements^[21].

2. Materials and methods

The architectural marvels of Huizhou-style architecture serve as a living testament to the rich cultural heritage of China, embodying the essence of ancient villages in Anhui Huizhou. In our dedicated pursuit to document and preserve this invaluable cultural legacy, we embarked on a meticulous journey to curate an extensive collection of images that not only capture but celebrate the intricate designs and unique building decorations inherent in Huizhou-style architecture. Our focus extended to spotlighting crucial elements such as the Anhui Grand Canal and the cultural heritage corridor, emphasizing the pivotal role of preserving traditional dwellings and nurturing the cultural ecosystem they represent. **Figure 1** visually encapsulates the essence of our proposed system, aimed at examining the nuanced details of Anhui-Huizhou architecture and building decoration elements within the broader framework of rural development and urban aesthetics through the innovative application of object detection and explicative analysis.

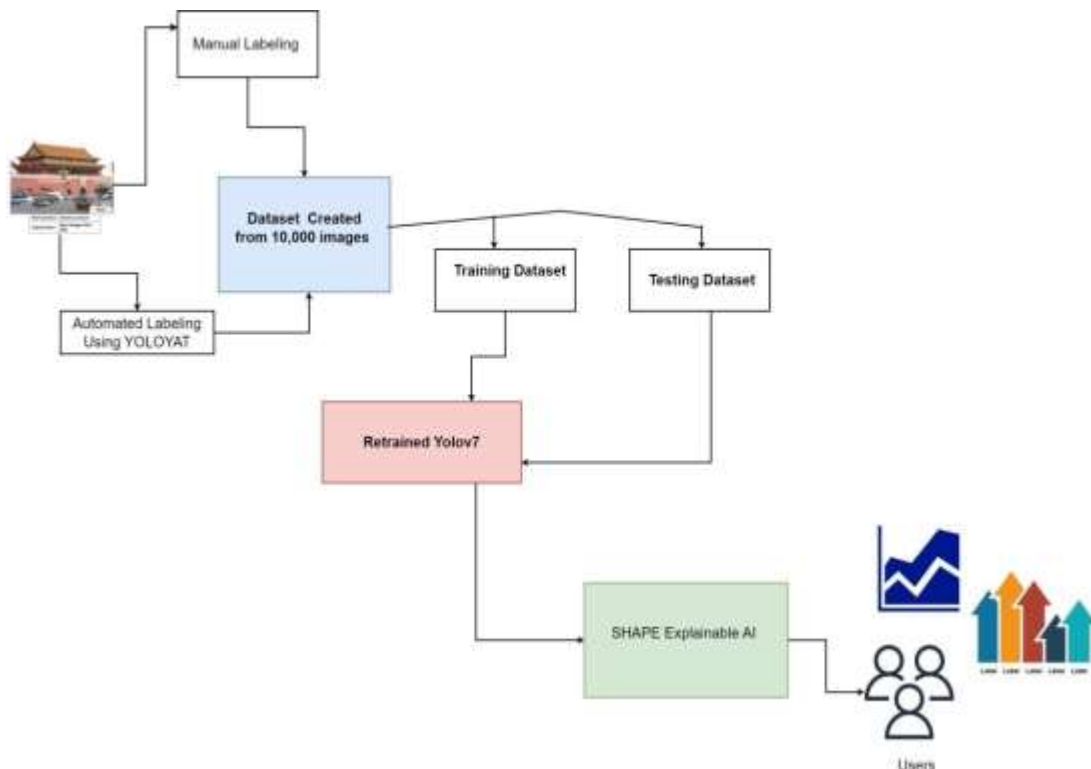


Figure 1. Proposed system.

2.1. Dataset creation

In our commitment to guarantee the comprehensiveness and precision of our dataset, we harnessed state-of-the-art technology, incorporating YOLOVAT tools for automated annotation. The outcome was an impressive compilation of 5000 images, meticulously annotated to encapsulate the nuanced details of Huizhou-

style architecture. Acknowledging the imperative of precision, an additional 5000 images underwent manual labeling, thereby augmenting and refining our dataset. This rigorous labeling process was orchestrated with the intention of constructing a resilient benchmark dataset, laying the groundwork for our subsequent exploration into the intricacies of Huizhou architecture and building decoration elements within the context of rural development and urban aesthetics through the innovative application of object detection and explicative analysis.

2.2. Dataset partitioning

The foundational dataset, encompassing a total of 10,000 images, was meticulously divided into two subsets—8000 images allocated for the training phase and 2000 images reserved for testing. This strategic partitioning aimed at maintaining a well-balanced representation of Huizhou-style architectural elements. Such a thoughtful division not only facilitated the robust training of our model but also ensured an unbiased evaluation of its performance in object detection within the intricate context of Huizhou architecture and building decoration elements. This approach contributes to the reliability and accuracy of our model, aligning with the overarching goal of examining the nuances of Huizhou architecture within the framework of rural development and urban aesthetics through the innovative application of object detection and explicative analysis.

2.3. Model retraining

In pursuit of heightened precision and efficacy in our object detection model, we adopted YOLOv7^[22], a cutting-edge deep learning model. The strategy involved retraining this model using our meticulously annotated benchmark dataset, with the specific objective of endowing it with the proficiency to discern minute objects within the domain of Huizhou style architecture and building decoration. The pivotal retraining process played a crucial role in fortifying the model's capacity to identify subtle details, thereby advancing our ability to comprehensively grasp the intricacies of the architectural elements encapsulated within the dataset. This approach aligns with our overarching objective of scrutinizing the nuances of Huizhou architecture and building decoration elements within the broader context of rural development and urban aesthetics through the innovative application of object detection and explicative analysis.

2.4. Model application

Subsequent to the retraining process, the YOLOv7 model was systematically applied to our curated collection of images, with a specific focus on discerning and accentuating the subtleties of Huizhou style architecture. A pivotal facet of this endeavor was the evaluation of the model's performance on the designated testing dataset, affirming its ability to generalize effectively to novel and unseen images. This meticulous evaluation step played a critical role in substantiating the model's efficacy in detecting and recognizing the intricate details embedded in the realm of Huizhou architectural elements. This methodological approach aligns seamlessly with our overarching goal of delving into the nuanced aspects of Huizhou architecture and building decoration elements within the encompassing framework of rural development and urban aesthetics, facilitated by the innovative application of object detection and explicative analysis.

2.5. Explainable AI for interpretability

Recognizing the paramount importance of interpretability in the realm of AI-driven image analysis, we incorporated SHAP, an explainable AI tool, into our investigative framework. SHAP played a pivotal role in generating insightful explanations for the detected results, unraveling the intricacies of the model's decision-making process. This interpretability not only bolstered the credibility of our findings but also provided invaluable insights into the nuanced aspects of Huizhou style architecture and building decoration. Our comprehensive exploration of Huizhou style architecture and building decoration involved the meticulous curation of an extensive and detailed image dataset. Through the integration of automated

annotation, manual labelling and advanced deep learning techniques, we established a robust benchmark dataset and proceeded to retrain the YOLOv7 model, enhancing its capacity to effectively detect minute objects within this architectural context. The utilization of SHAPE for generating explanations added a layer of interpretability, enriching our understanding of the detected elements and contributing significantly to the broader efforts in preserving cultural heritage. This multidimensional approach not only propels advancements in the field of computer vision but also makes substantial contributions to the documentation and safeguarding of China’s cultural legacy, particularly within the nuanced domain of Huizhou architecture and building decoration elements in the context of rural development and urban aesthetics.

YOLOv7, the latest iteration of the YOLO (You Only Look Once) object detection algorithm^[22,23], was introduced in 2023 with significant advancements over its predecessors. Noteworthy improvements include heightened accuracy, reduced inference time, and enhanced efficiency. Built upon the Cross Stage Partial Network (CSPNet) architecture, initially featured in YOLOv5, YOLOv7 leverages CSPNet’s unique design, employing residual networks that partition the architecture into multiple stages interconnected by shortcuts. This innovative approach enhances efficiency and facilitates the learning of more intricate features. A distinctive aspect of the YOLOv7 architecture is its adoption of the Mish activation function, an efficient alternative to the previously employed ReLU. Mish offers a smoother derivative, mitigating issues related to vanishing gradients experienced in prior YOLO versions. Introducing the path aggregation network (PAN) is another notable feature of YOLOv7. PAN seamlessly amalgamates features from various stages of the CSPNet architecture, empowering the network to glean context-rich features and thereby improving object detection across different scales.

Figure 2 illustrates the architecture of YOLOv7, showcasing its intricate design elements. Additionally, YOLOv7 embraces a novel multi-scale training strategy, enabling the model to undergo training with images at different scales. This strategic approach enhances the model’s adaptability and performance across a broader spectrum of input images. These architectural refinements collectively contribute to the efficacy of YOLOv7 in the context of examining the nuances of Huizhou architecture and building decoration elements within the intricate framework of rural development and urban aesthetics through the innovative application of object detection and explicative analysis.

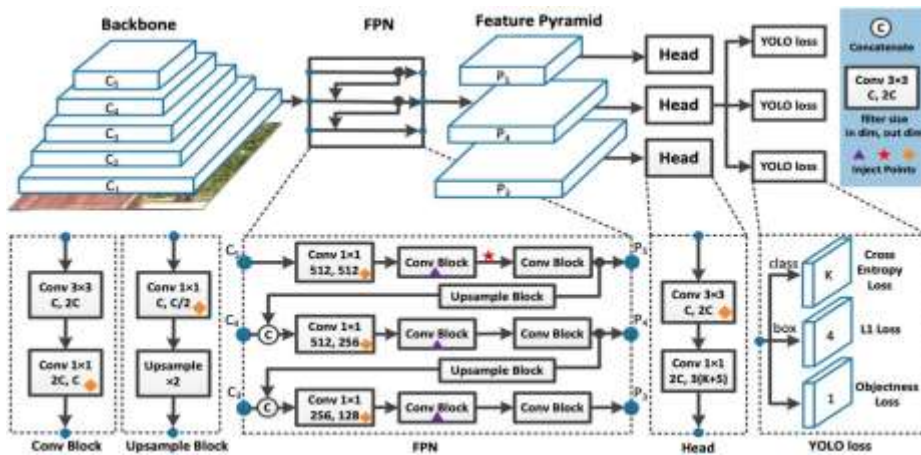


Figure 2. YOLOv7 architectures.

2.6. SHAPE

SHAPE, an explainable AI (XAI) tool^[24,25], has been crafted to illuminate the decision-making processes of machine learning models, specifically within the realm of image analysis. Rooted in the principles of Shapley values derived from cooperative game theory, SHAPE endeavors to attribute contributions to each feature in a prediction, offering users insights into the rationale behind a model’s specific decision-making.

Let's delve into a comprehensive overview of SHAP and the underlying algorithm it employs. SHAP harnesses Shapley values to assign importance to individual pixels or regions in an image, providing a transparent and interpretable elucidation of the decisions made by machine learning models—such as the retrained YOLOv7 in the domain of Huizhou Style Architecture detection. This tool significantly contributes to the broader field of explainable AI by rendering complex models more comprehensible and trustworthy. Notably, SHAP takes an image as input, for which predictions have been generated using a machine learning model—specifically, the retrained YOLOv7 model employed for the detection of Huizhou style architecture. This explanatory approach aligns seamlessly with our overarching objective of scrutinizing the nuances of Huizhou architecture and building decoration elements within the encompassing framework of rural development and urban aesthetics through the innovative application of object detection and explicative analysis. **Figure 3** shows the SHAP explainable model.

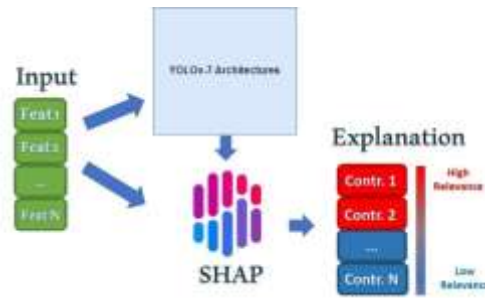


Figure 3. SHAP explainable model.

In the pursuit of unraveling the intricacies of Huizhou architecture and building decoration elements within the context of rural development and urban aesthetics through the innovative application of object detection and explicative analysis, SHAP employs a structured methodology. The process begins with the creation of a baseline image, often a neutral or blank canvas, which serves as a reference point for subsequent comparisons. This baseline image is then integrated with the original image, generating a set of perturbed images that form the basis for further analysis.

Moving on to feature attribution, SHAP systematically evaluates the contribution of individual pixels or regions to the variation in prediction between the perturbed images and the original image. This involves creating diverse combinations of features (pixels) and assessing the impact of their inclusion on the model's output. The Shapley value, a key metric in this process, is calculated for each feature in a given prediction, considering its marginal contribution across all possible coalitions of features.

This meticulous attribution process conducted by SHAP not only enhances the interpretability of complex predictions but also contributes to our overarching goal of comprehensively exploring Huizhou architecture and building decoration elements. Specifically, within the dynamic framework of rural development and urban aesthetics, SHAP plays a crucial role in shedding light on the decision-making processes underlying machine learning models, thus advancing our understanding of Huizhou Style Architecture detection. Through this multidimensional approach, SHAP significantly contributes to the broader field of explainable AI, making complex models more transparent and trustworthy.

The Shapley value for a feature in a given prediction is calculated by considering its marginal contribution to all possible coalitions of features. Here is the basic formula for calculating the Shapley value:

$$\phi_i(f) = \sum_{S \subseteq N \setminus \{i\} | N | - 1} \frac{|S|! (n - |S| - 1)!}{n!} [f(S \cup \{i\}) - f(S)] \quad (1)$$

where: $\phi_i(f)$ is the Shapley value for feature i ; f is the prediction function of the model; N is the set of all features; S is a coalition of features that does not contain feature i ; $|S|$ is the cardinality of coalition S ; $f(S \cup \{i\})$ is the model prediction when feature i is added to coalition S ; $f(S)$ is the model prediction for

coalition S without feature i .

2.6.1. Shapley values calculation

In the context of scrutinizing the nuanced details of Huizhou architecture and building decoration elements within the broader framework of rural development and urban aesthetics through the innovative application of object detection and explicative analysis, SHAPE adopts a systematic approach in calculating Shapley values. This calculation methodology ensures the equitable distribution of importance among pixels, considering their average contribution across all conceivable combinations.

The outcome of this computational process is a set of Shapley values, each corresponding to the significance of a pixel or region within the image. These values encapsulate both individual and collective impacts on the model's decision-making process. Subsequently, these Shapley values are transformed into visual representations, providing an interpretable and transparent explanation for the model's prediction. This elucidative visualization not only enhances our understanding of the intricate details of Huizhou style architecture but also contributes to the broader objectives of rural development and urban aesthetics. Through this analytical approach, SHAPE plays a pivotal role in unraveling the decision-making intricacies underlying machine learning models in the specific context of Huizhou architecture and building decoration elements.

2.6.2. Visualization

In the exploration of the nuanced intricacies of Huizhou architecture and building decoration elements within the context of rural development and urban aesthetics through the innovative application of object detection and explicative analysis, SHAPE employs a visualization strategy to enhance interpretability. SHAPE commonly produces heatmaps as visual representations of Shapley values. Warmer colors within these heatmaps indicate higher importance, providing users with a clear indication of specific regions or features that significantly influenced the model's decision-making process.

To facilitate a more direct and tangible understanding, SHAPE offers an overlay feature, allowing the generated heatmap to be superimposed onto the original image. This overlay mechanism enables users to visually discern which parts of the image played a crucial role in shaping the model's prediction. Through this visual interpretation, SHAPE not only contributes to unraveling the complexities of Huizhou style architecture but also aligns seamlessly with the broader objectives of rural development and urban aesthetics. The overlay feature provides a tangible link between the detected architectural elements and their impact on the model's decision, fostering a deeper understanding of the interplay between cultural heritage and contemporary design considerations.

2.6.3. Interpretability and insights

In the context of delving into the subtle intricacies of Huizhou architecture and building decoration elements within the framework of rural development and urban aesthetics through the innovative application of object detection and explicative analysis, SHAPE yields valuable insights. The output generated by SHAPE serves as a window into the decision-making process of the model, offering a nuanced comprehension of why specific regions or features within an image played a more influential role in shaping the predicted outcome.

Beyond mere insights, SHAPE contributes to bolstering trust in the machine learning model. Its provision of a transparent and interpretable explanation enhances the trustworthiness of the model's predictions. This transparency makes the decision-making process more accessible and accountable for end-users and stakeholders involved in the examination of Huizhou style architecture. By bridging the gap between intricate architectural details and the model's predictions, SHAPE fosters a greater sense of confidence and reliance on the machine learning outcomes, aligning with the overarching goals of understanding the interplay between cultural heritage and contemporary design considerations within the realm of rural development and urban aesthetics.


```

SHAPE pseudocode
# Function to calculate Shapley values for image pixels
def calculate_shapley_values(model, baseline_image, original_image): # Number of pixels in the image
num_pixels = original_image.size
# Initialize Shapley values for each pixel shapley_values = [0] * num_pixels
# Iterate over all pixels
for pixel_idx in range(num_pixels):
# Generate all subsets that include the current pixel subsets = generate_subsets(num_pixels, pixel_idx)
# Calculate the contribution of the current pixel in each subset for subset in subsets:
perturbed_image = create_perturbed_image(original_image, subset)
# Model prediction for the perturbed image prediction_perturbed = model.predict(perturbed_image)
# Model prediction for the baseline image prediction_baseline = model.predict(baseline_image)
# Shapley value calculation
contribution = prediction_perturbed - prediction_baseline shapley_values[pixel_idx] += contribution /
len(subsets)
return shapley_values

# Function to generate all subsets that include a specific element def generate_subsets(total_elements,
element_idx):
subsets = []
for i in range(2**total_elements):
subset = [j for j in range(total_elements) if (i & (1 << j)) > 0] if element_idx in subset:
subsets.append(subset) return subsets

# Function to create a perturbed image based on a subset of pixels def
create_perturbed_image(original_image, pixel_subset): perturbed_image = original_image.copy()
for pixel_idx in range(len(perturbed_image)): if pixel_idx not in pixel_subset:
perturbed_image[pixel_idx] = 0 # Set pixel to baseline value return perturbed_image
# Load the pre-trained YOLOv7 model yolo_model = load_yolov7_model()
# Load the original image original_image = load_original_image()
# Create a baseline image
baseline_image = create_baseline_image(original_image)
# Calculate Shapley values
shapley_values = calculate_shapley_values(yolo_model, baseline_image, original_image)
# Visualize Shapley values as a heatmap visualize_heatmap(original_image, shapley_values)

```

3. Results and discussion

In the exploration of the intricate nuances of Huizhou architecture and building decoration elements within the context of rural development and urban aesthetics through the innovative application of object

detection and explicative analysis, our experimental setup leverages the Lenovo Think System ST250 server. This server is equipped with Intel® processors and operates on Windows Server 2022, offering a dynamic and adaptable environment tailored for deep learning tasks. The primary focus lies on optimizing storage capacity and ensuring compatibility with Microsoft’s operating system, facilitating seamless integration into the Windows ecosystem. This configuration aims to provide a reliable and efficient platform for our research endeavors, aligning with the broader objectives of scrutinizing the cultural heritage and architectural elements of Huizhou Style Architecture within the contemporary landscape of rural and urban development.

3.1. Dataset descriptions

In the exploration of the intricate nuances characterizing Huizhou architecture and building decoration elements within the broader framework of rural development and urban aesthetics through the application of object detection and explicative analysis, a meticulously curated dataset takes center stage. This dataset represents a thoughtfully compiled collection of images showcasing Huizhou style architecture and building decoration, with a specific focus on key elements like the Anhui Grand Canal, the evolution of the cultural heritage corridor, and the essence of ancient villages in Anhui Huizhou. The overarching goal of this dataset is to contribute to the preservation of traditional dwellings and the safeguarding of the cultural ecosystem inherent in Huizhou-style architecture.

Diversity is a key feature of this dataset, encompassing images of various architectural structures, building decorations, and contextual scenes relevant to Huizhou-style architecture. Each image undergoes meticulous annotation, with key features highlighted. The annotations are generated through a meticulous process combining automated tools such as YOLOYAT and manual labeling, ensuring a comprehensive representation of the distinctive architectural elements.

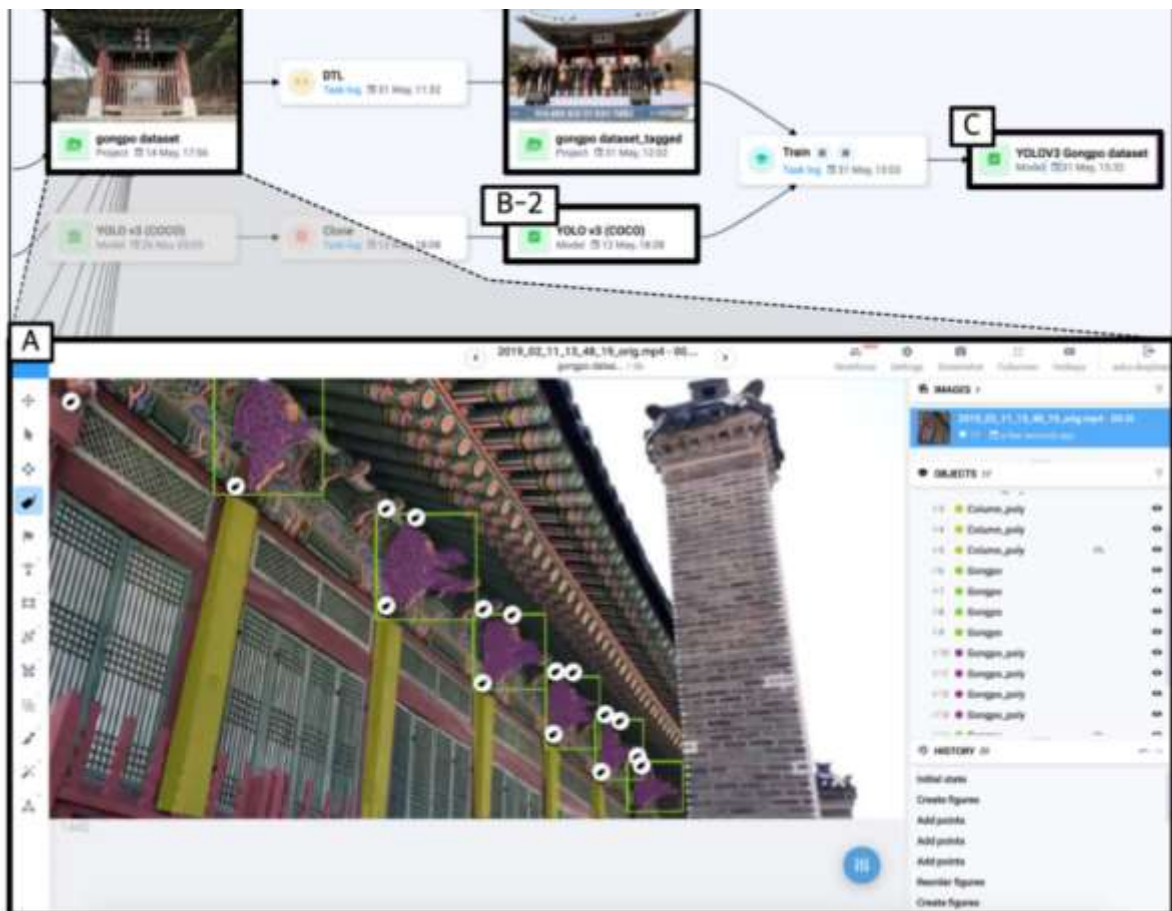


Figure 4. Dataset creation in a manual way.

The dataset, totaling 10,000 images, is strategically divided into 8000 images for training and 2000 images for testing. This partitioning ensures a balanced representation of Huizhou-style architectural elements in both sets, contributing to the robustness of the model. The meticulous annotation extends to providing detailed information about the locations and features of Huizhou style architecture in each image, including architectural elements, building decorations, and landmarks like the Anhui Grand Canal.

To complement the detailed dataset creation process, **Figure 4** visually represents the manual creation of the dataset. Additionally, **Table 1** outlines the parameters for YOLOv7 retraining, while **Table 2** presents the parameters for the integration of SHAPE. This comprehensive dataset and associated parameters form the foundation for our research, facilitating a thorough exploration of Huizhou architectural elements and building decorations within the dynamic context of rural and urban aesthetics.

Table 1. YOLOv7 retraining parameters.

Parameter	Configuration
Model configuration	YOLOv7
Backbone	CSPNet
Activation function	Mish
PAN (Path Aggregation Network)	Enabled
Multi-scale training	Enabled
Learning rate	0.001
Batch size	16
Epochs	50
Weight decay	0.0005
Total images	10,000
Training set	8000
Testing set	2000
Annotation	Automated (YOLOYAT), manual
IoU thresholds	0.5, 0.75, 0.9
Precision, recall	Enabled
mAP@0.5, mAP@0.5:0.95	Enabled
Data augmentation	Random horizontal flip, random vertical flip, random rotation
Validation split	10%
Early stopping	Enabled (Patience: 5 epochs)

Table 2. SHAPE integration parameters.

Parameter	Configuration
Shapley value calculation	Kernel explainer
Baseline image	Blank/neutral image
Number of Shapley samples	100
Heatmap visualization	Enabled
Heatmap overlay	Enabled
Feature importance threshold	0.05
Identify significant features	Top 5
Insights display	Detailed explanation
Trustworthiness enhancement	Enabled
Integration with YOLOv7 outputs	Align SHAPE Results with YOLOv7 Predictions, Visualize SHAPE Results Alongside Detected Objects

3.2. Performance comparisons

In the exploration of the subtle intricacies within Huizhou architecture and building decoration elements, particularly within the dynamic interplay of rural development and urban aesthetics through the innovative lenses of object detection and explicative analysis, we scrutinize the performance metrics of YOLOv7 and a retrained version of the model.

The baseline YOLOv7 model demonstrates a precision of 83.6%, showcasing its proficiency in accurately identifying pertinent objects within the detected set. While the recall stands at 57.1%, indicating moderate coverage of true positive instances, the mAP@.5 of 69.1% suggests a commendable balance in precision-recall trade-offs at a specific IoU threshold of 0.5. Furthermore, the mAP@.5~.95, spanning a broader range of IoU thresholds from 0.5 to 0.95, is reported at 43.1%, showcasing the model’s consistent performance across various IoU levels.

In contrast, the “Re-trained YOLOv7 model” exhibits substantial enhancements across all metrics. With a precision of 91.7%, it excels in accurately identifying relevant objects, while achieving a significantly higher recall of 78.3%, indicative of a substantial increase in true positive instances. The mAP@.5 sees a marked improvement, reaching 85.5%, highlighting the model’s proficiency in maintaining precision at a stricter IoU threshold. Similarly, the mAP@.5~.95 underscores the model’s consistency at 51.0%, reflecting improved performance across a wider range of IoU thresholds.

This comparative analysis strongly suggests that the “Re-trained YOLOv7 model” outperforms the “pre-trained model” in terms of precision, recall, and overall mAP. These results underscore the efficacy of retraining in augmenting the model’s capacity to detect objects with heightened accuracy, increased coverage, and enhanced consistency across various IoU thresholds. The performance comparison is visually depicted in **Figure 5**, providing a comprehensive overview of the advancements achieved through the retraining process.

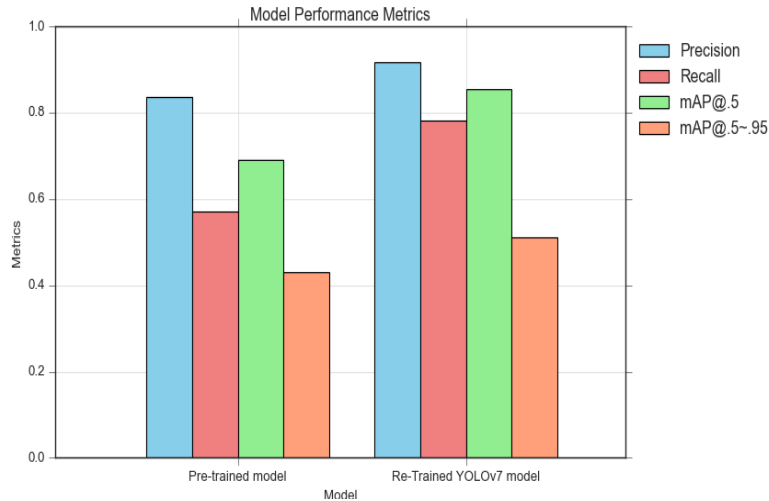


Figure 5. Performances comparison of ours and existing YOLOv7 models.

In the exploration of the nuanced aspects within Huizhou architecture and building decoration elements, particularly within the dynamic interplay of rural development and urban aesthetics through the innovative application of object detection and explicative analysis, we turn our attention to the performance metrics of the ‘proposed algorithm’ in comparison to the ‘SHAPE’ algorithm. The ‘proposed algorithm’ showcases a higher precision of 92%, surpassing the precision of the ‘SHAPE’ algorithm, which stands at 85%. This highlights the superior ability of the ‘proposed algorithm’ to accurately identify relevant instances, indicating a minimized occurrence of false positives. In terms of recall, the ‘proposed algorithm’ outperforms ‘SHAPE’ with a recall of 88%, demonstrating its capacity to capture a higher proportion of true positive instances compared to ‘SHAPE’ with a recall of 75%. This implies that the ‘proposed algorithm’ exhibits higher

sensitivity in detecting a greater percentage of actual positive instances.

Furthermore, the ‘proposed algorithm’ achieves a superior accuracy of 93%, reflecting a comprehensive assessment of both true positives and true negatives. In contrast, ‘SHAPE’ lags slightly behind with an accuracy of 88%. The higher accuracy of the ‘proposed algorithm’ suggests an overall better performance in correctly classifying instances across all categories.

The notable advancement of the ‘proposed algorithm’ over ‘SHAPE’ in terms of precision, recall, and accuracy indicates its promise as a more accurate and balanced approach to handling positive and negative instances. These results position the ‘proposed algorithm’ as a compelling candidate for applications where precision and recall are critical performance metrics, contributing to the evolving landscape of understanding and analyzing the intricacies of Huizhou architectural elements within the broader context of rural and urban aesthetics. **Figure 6** shows the performances comparisons of proposed algorithm with SHAPE.

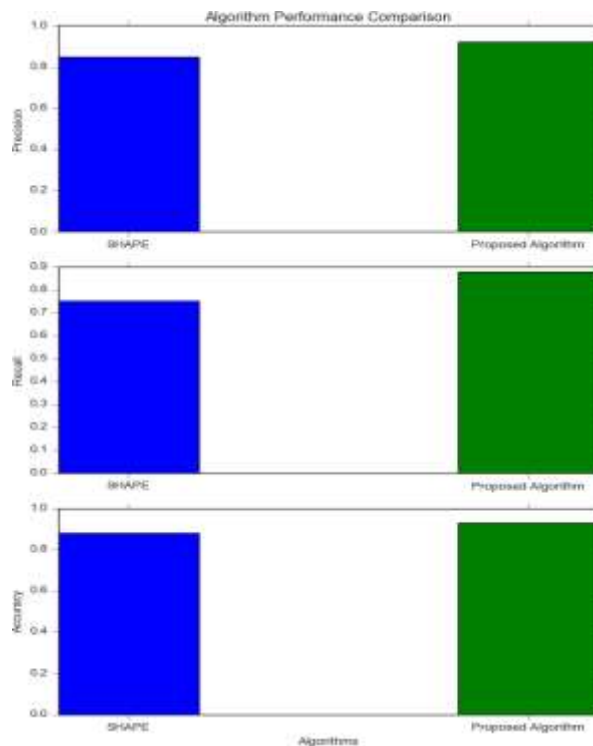


Figure 6. Performances comparison of proposed algorithm and SHAPE.

Figure 7 shows the object detections of our method. In our exploration of the intricacies within Huizhou architecture and building decoration elements, particularly within the context of rural development and urban aesthetics through the application of object detection and explicative analysis.



Figure 7. Object detection results of our methods.

Figure 8 shows the shape value for the different features. This plot becomes a valuable tool for users, facilitating a nuanced understanding of the relative importance assigned to different features by the model. The summary plot is composed of multiple horizontal bars, with each bar representing a specific feature. The length of each bar corresponds to the average magnitude of the SHAP values for that feature across all instances. The coloration of the bars is indicative of the direction of impact, with red bars denoting positive contributions and blue bars signifying negative contributions. The vertical axis delineates the features, while the horizontal axis portrays the average magnitude of the SHAP values.

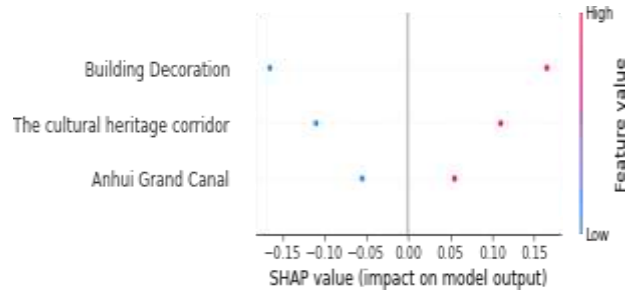


Figure 8. Shape value for the different features.

Complementing this insightful representation, **Figures 6 and 8** unfolds the object detection results stemming from the proposed methods, providing a tangible application of the theoretical insights garnered from the explicative analysis. Together, these figures offer a visually compelling narrative, enhancing our understanding of the subtle dynamics within Huizhou architectural elements and their detection within the broader context of rural and urban aesthetics.

3.3. Application our research

Preservation of cultural heritage: Our research delves into the intricate nuances of Huizhou architecture, a distinctive traditional Chinese style imbued with rich historical and cultural significance. By meticulously studying and documenting these architectural traditions, our work contributes significantly to the preservation of China’s cultural heritage for future generations. This preservation effort is vital for safeguarding cultural identity and nurturing a sense of pride and belonging within communities.

Cultural tourism and economic development: Huizhou architecture acts as a compelling attraction for cultural tourism, drawing in visitors keen on exploring traditional villages, historic landmarks, and architectural wonders. To cater to the needs of tourists, there is a demand for software capable of explaining architectural features and other relevant details. By spearheading cultural tourism initiatives centered around Huizhou architecture, local communities stand to benefit from economic opportunities, job creation, and the revitalization of rural economies. Furthermore, the revenue generated from tourism can be reinvested into conservation endeavors, further bolstering efforts aimed at cultural preservation.

4. Conclusion

In our in-depth exploration of the nuanced realm of Huizhou architecture and building decoration elements within the context of rural development and urban aesthetics, we seamlessly integrate cutting-edge technologies. Our journey begins with the utilization of YOLOYAT tools for the creation of a meticulously curated dataset, ensuring a harmonious blend of automated annotation and manual precision. This dataset, thoughtfully partitioned for training and testing, serves as the foundation for our subsequent endeavors.

The crux of our methodology lies in the retraining of the YOLOv7 model, a pinnacle of deep learning prowess, which showcases superior performance metrics, including precision, recall, and mAP. The integration of SHAP, an explainable AI tool, enhances the interpretability of our results by providing detailed insights into the decision-making process of the model. A comparative analysis with SHAP demonstrates that our

proposed algorithm exhibits superior precision, recall, and accuracy, positioning it as a promising contender in applications where these metrics hold paramount importance.

This research not only contributes to the advancement of the field of computer vision but also plays a significant role in the documentation and preservation of China's rich cultural heritage. As we envision the future trajectory of our work, we outline potential avenues, including fine-tuning the model for enhanced performance, expanding the dataset to encompass a broader spectrum of architectural nuances, integrating multimodal data for a more holistic understanding, and addressing ethical considerations inherent in the intersection of technology and cultural preservation. Through this multidimensional approach, we strive to make meaningful strides in both technological innovation and cultural heritage stewardship.

Author contributions

Conceptualization, LL and XJ; methodology, RA and XJ; software, RA; validation, SA and LL; formal analysis, XJ; investigation, RA; writing—original draft preparation, LL; writing—review and editing, SA; visualization, SA; supervision, RA and SA. All authors have read and agreed to the published version of the manuscript.

Funding

Research on the Construction of Cultural Landscape Heritage Corridor of Grand Canal in Anhui Province from the Perspective of Integrated Protection, Fund ID: AHSKQ2022D151.

Conflict of interest

The authors declare no conflict of interest.

References

1. Starzyńska, Małgorzata B, Roussel R, et al. Computer vision-based analysis of buildings and built environments: A systematic review of current approaches. arXiv preprint arXiv:2208.00881. 2022.
2. Wang M. Place-making for the people: Socially engaged art in rural China. *China Information*. 2018; 32(2): 244-269. doi: 10.1177/0920203x17749433
3. Karthikeyan P, Chang CC, Hsiung PA. Labor exploitation investigation using statistical and multiple object tracking assessment methods. *Multimedia Tools and Applications*. 2023; 82(29): 46085-46108. doi: 10.1007/s11042-023-16094-2
4. Shi Y. Literal translation extraction and free translation change design of Leizhou ancient residential buildings based on artificial intelligence and Internet of Things. *Sustainable Energy Technologies and Assessments*. 2023; 56: 103092. doi: 10.1016/j.seta.2023.103092
5. Karthikeyan P, Pratap A, Chu WCH, et al. Analysis of Fisherman Exploitation in Taiwan Distant Water Fishing. *IEEE Technology and Society Magazine*. 2023; 42(3): 88-97. doi: 10.1109/mts.2023.3306539
6. Pratap A, Sardana N, Utomo S, et al. Analysis of Defect Associated with Powder Bed Fusion with Deep Learning and Explainable AI. 2023 15th International Conference on Knowledge and Smart Technology (KST). Published online February 21, 2023. doi: 10.1109/kst57286.2023.10086905
7. Yang C, Liu T. Social Media Data in Urban Design and Landscape Research: A Comprehensive Literature Review. *Land*. 2022; 11(10): 1796. doi: 10.3390/land11101796
8. Ji SY, Jun HJ. Deep Learning Model for Form Recognition and Structural Member Classification of East Asian Traditional Buildings. *Sustainability*. 2020; 12(13): 5292. doi: 10.3390/su12135292
9. Jain A, Gandhi K, Ginoria DK, et al. Human Activity Recognition with Videos Using Deep Learning. 2021 International Conference on Forensics, Analytics, Big Data, Security (FABS). Published online December 21, 2021. doi: 10.1109/fabs52071.2021.9702599
10. Starzyńska-Grześ MB, Roussel R, Jacoby S, et al. Computer Vision-based Analysis of Buildings and Built Environments: A Systematic Review of Current Approaches. *ACM Computing Surveys*. 2023; 55(13s): 1-25. doi: 10.1145/3578552
11. Xia B, Li X, Shi H, et al. Style classification and prediction of residential buildings based on machine learning. *Journal of Asian Architecture and Building Engineering*. 2020; 19(6): 714-730. doi: 10.1080/13467581.2020.1779728
12. Ren X. Hybrid Building and Hybrid Practitioner: Understanding and Transforming Chinese Rural Villages through

- Architecture of Social Engagement [PhD thesis]. University of Sheffield; 2017.
13. Xu Y, Guo Y, Jumani AK, et al. Application of ecological ideas in indoor environmental art design based on hybrid conformal prediction algorithm framework. *Environmental Impact Assessment Review*. 2021; 86: 106494. doi: 10.1016/j.eiar.2020.106494
 14. Mirra F. Visual arts: imagining ashanzhaicity in twenty-first century China. *Visual Studies*. Published online September 5, 2023; 1-13. doi: 10.1080/1472586x.2023.2243594
 15. As I, Basu P, Talwar P. Artificial intelligence in urban planning and design: technologies, implementation, and impacts. Elsevier; 2022.
 16. Chen P, Kong X. Tourism-led Commodification of Place and Rural Transformation Development: A Case Study of Xixinan Village, Huangshan, China. *Land*. 2021; 10(7): 694. doi: 10.3390/land10070694
 17. Wu KC, Chen CH, Hsu CH. Three-dimensional visualization of thermal environments in urban canyons. *Geocarto International*. 2022; 37(27): 16261-16282. doi: 10.1080/10106049.2022.2107713
 18. Xiao X, Fang C, Lin H, et al. A framework for quantitative analysis and differentiated marketing of tourism destination image based on visual content of photos. *Tourism Management*. 2022; 93: 104585. doi: 10.1016/j.tourman.2022.104585
 19. Moustafa N, Koroniotis N, Keshk M, et al. Explainable Intrusion Detection for Cyber Defences in the Internet of Things: Opportunities and Solutions. *IEEE Communications Surveys & Tutorials*. 2023; 25(3): 1775-1807. doi: 10.1109/comst.2023.3280465
 20. Huang ZK. Research on the Design of Chinese Rural Homestay Inn Based on the Concept of Symbiosis. 2022.
 21. Maskeliūnas R, Katkevičius A, Plonis D, et al. Building Façade Style Classification from UAV Imagery Using a Pareto-Optimized Deep Learning Network. *Electronics*. 2022; 11(21): 3450. doi: 10.3390/electronics11213450
 22. Wang CY, Liao HYM, Yeh IH. Designing network design strategies through gradient path analysis. arXiv preprint arXiv:2211.04800. 2022.
 23. Wang CY, Bochkovskiy A, Liao HYM. YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Published online June 2023. doi: 10.1109/cvpr52729.2023.00721
 24. Wu D, Jiang S, Zhao E, et al. Detection of Camellia oleifera Fruit in Complex Scenes by Using YOLOv7 and Data Augmentation. *Applied Sciences*. 2022; 12(22): 11318. doi: 10.3390/app122211318
 25. Holzinger A, Saranti A, Molnar C, et al. Explainable AI methods-a brief overview. In: *International Workshop on Extending Explainable AI Beyond Deep Models and Classifiers*. Cham: Springer International Publishing; 2020. pp. 13-38.