



## Article

# Central Courtyard Feature Extraction in Remote Sensing Aerial Images Using Deep Learning: A Case-Study of Iran

Hadi Yazdi <sup>1,2</sup> , Ilija Vukorep <sup>2</sup> , Marzena Banach <sup>1</sup>, Sajad Moazen <sup>3</sup> , Adam Nadolny <sup>1</sup> , Rolf Starke <sup>2</sup> and Hassan Bazazzadeh <sup>1,\*</sup>

<sup>1</sup> Faculty of Architecture, Poznan University of Technology, 61-131 Poznan, Poland; hadi.yazdi@b-tu.de (H.Y.); marzena.banach@put.poznan.pl (M.B.); adam.nadolny@put.poznan.pl (A.N.)

<sup>2</sup> Faculty of Architecture, Civil Engineering and Urban Planning, Brandenburg University of Technology, 03046 Cottbus, Germany; ilija.vukorep@b-tu.de (I.V.); Rolf.Starke@b-tu.de (R.S.)

<sup>3</sup> School of Architecture and Environmental Design, Iran University of Science and Technology, Tehran 16846-13114, Iran; sajadmoazen@iust.ac.ir

\* Correspondence: hassan.bazazzadeh@doctorate.put.poznan.pl

**Abstract:** Central courtyards are primary components of vernacular architecture in Iran. The directions, dimensions, ratios, and other characteristics of central courtyards are critical for studying historical passive cooling and heating solutions. Several studies on central courtyards have compared their features in different cities and climatic zones in Iran. In this study, deep learning methods for object detection and image segmentation are applied to aerial images, to extract the features of central courtyards. The case study explores aerial images of nine historical cities in Bsk, Bsh, Bwk, and Bwh Köppen climate zones. Furthermore, these features were gathered in an extensive dataset, with 26,437 samples and 76 geometric and climatic features. Additionally, the data analysis methods reveal significant correlations between various features, such as the length and width of courtyards. In all cities, the correlation coefficient between these two characteristics is approximately +0.88. Numerous mathematical equations are generated for each city and climate zone by fitting the linear regression model to these data in different cities and climate zones. These equations can be used as proposed design models to assist designers and researchers in predicting and locating the best courtyard houses in Iran's historical regions.

**Keywords:** central courtyard; deep learning; features extraction; data analysis; Iran



**Citation:** Yazdi, H.; Vukorep, I.; Banach-Ziaja, M.; Moazen, S.; Nadolny, A.; Starke, R.; Bazazzadeh, H. Central Courtyard Feature Extraction in Remote Sensing Aerial Images Using Deep Learning: A Case-Study of Iran. *Remote Sens.* **2021**, *13*, 4843. <https://doi.org/10.3390/rs13234843>

Academic Editors: Cettina Santagati, Ilaria Trizio and Belen Riveiro

Received: 6 October 2021

Accepted: 15 November 2021

Published: 29 November 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

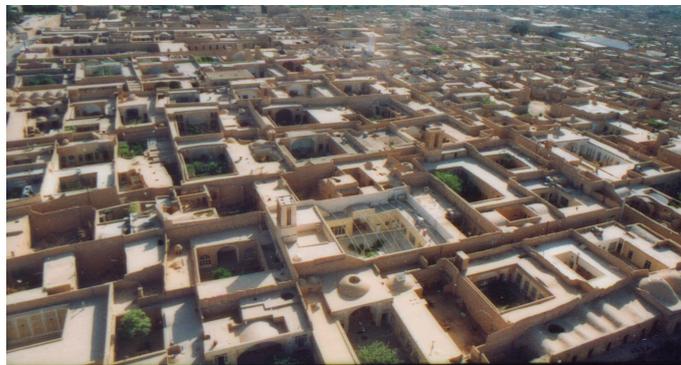
## 1. Introduction

### 1.1. Historical Architecture of Iran

Vernacular architecture, which has developed throughout time, uses a variety of novel techniques and technologies to address a variety of contextual requirements in order to improve sustainability [1,2] and urban identity [3]. Moreover, historic and vernacular architecture as an attraction of cities [4], it can help the economic prosperity [5]. Iran's vernacular architecture has perfected the skill of context adaptation by developing distinct structures in the country's many areas. Despite their fundamental differences, vernacular designs of these regions have a common philosophical underpinning: sustained contextual adaptability [6,7]. Indeed, by definition, a vernacular structure is constructed by locals, employing traditional methods and locally accessible materials to suit domestic ways of living [8], which is consistent with what is currently known as sustainability. That is why it is highly critical to protect these structures instead of reconstruct them [9]. The house's cooling and heating systems are primarily passive [10] in regard to fulfilling varying demands [11], which results in varying architectural approaches in vernacular structures. This strategies are very important in the sustainability of this style of architecture as studies have shown that the rate of energy consumption of buildings may increase in the future due to future uncertainties [12,13] and the design should be optimized in all aspects [14].

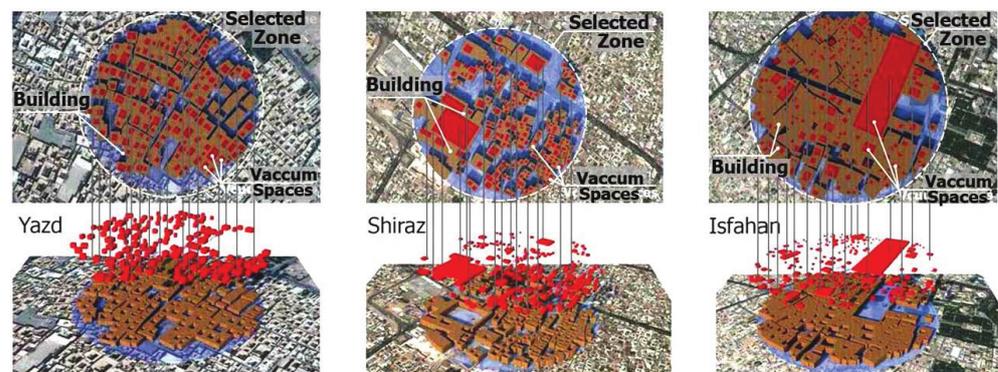
Iranian architects have developed various solutions and methods from natural characteristics to suit the country's diverse climatic conditions [15]. In traditional Iranian architecture, cost efficiency was among the most critical factors to be considered. Therefore, materials used in desert architecture must be readily accessible, cheap, and economical. Additionally, vernacular architects used locally accessible materials to construct a haven from atmospheric pollutants [16].

A great example of sustainability in Iranian architecture can be found in Yazd, which has the most significant surviving historic urban fabric in the country. Due to Yazd's hot and dry climate, providing occupants with high living standards was a real challenge for architects [17]. The requirement to protect people from the summer heat and scorching sun has significantly impacted the design methods used to construct shade zones. Thus, the urban design was dense, to provide maximum protection from inclement weather [18,19], and central courtyards were used to meet daylight requirements and fit the design with the climate (Figure 1).



**Figure 1.** Urban configuration of Yazd.

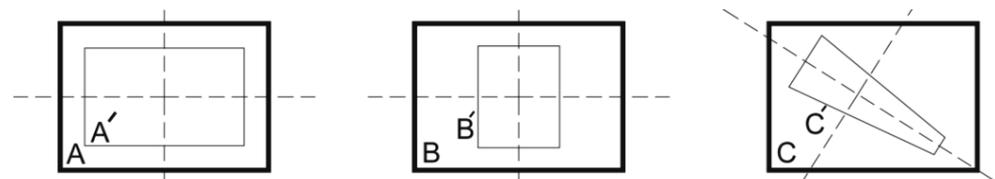
As an indistinguishable element of houses in the central plateau of Iran, the central courtyard played a structural role at the macro, middle, and micro scales in municipalities, segments, and structures, and similarly, as a repeated element, presented the vacuum concept as a fundamental component of a coherent urban development model. Central courtyards (vacuum spatial grid) serve as an organizer in the microsystem of structures, particularly small classical houses in Iran (Figure 2). Since spaces are organized around it, and when the courtyard is excluded, the house's spatial organization is disrupted [20].



**Figure 2.** The connection between vacuum spaces and urban designs in Yazd, Shiraz, and Esfahan [20].

Houses with courtyards have been utilized and improved for an extended period because they are advantageous for building in hot-arid climates. Most ancient civilizations used this technique in central Iran and several other arid regions of the Middle East. In hot-arid regions and deserts, central yards effectively protect people from severe environmental conditions. In desert regions, rooms that face the yard are often sheltered from severe heat,

winter cold, windstorms, and sand [19,21,22]. As part of the house's structure, the courtyard has a consistent structure and will improve and complement space arrangement. Understanding the relationship between components in the organizational structures of Iranian homes is critical to comprehend their interactions. Thus, the connection between the components and the whole in any spatial organization system is determined by various variables, including form, material, size, color, function, and content. Fundamentally, in the relationship between parts and a whole, if one part differs from the others, in one or more characteristics, it represents a distinct component that exists independently of the whole and functions within an organizational system. The component's form is the most significant predictor of a reciprocal connection between it and the whole (Figure 3).



**Figure 3.** Relationships between components and the formation of a cohesive whole [20].

### 1.2. Deep Learning, Architecture, and Planning

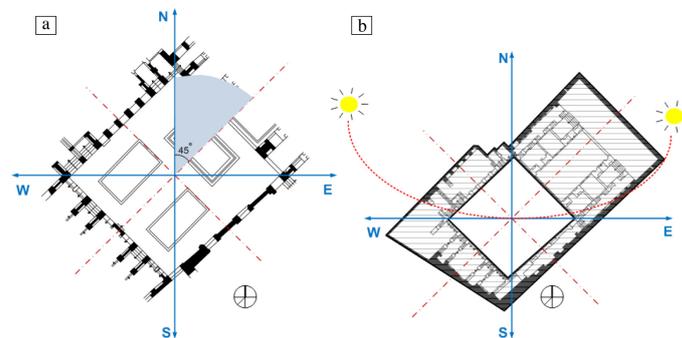
Until recently, computational techniques for structural analysis, design, and development have been almost unmatched by methods that analyze the current environment visually, and convert it to digital data and in-depth information. A few years ago, the field of computer vision made significant strides in developing programs for identifying and classifying existing components for various purposes. AlexNet, a convolutional neural network, may be seen as a forerunner to this 2012 advancement [23]. Airborne and satellite imaging remote sensing has also become a critical element of classical architecture and archaeological research, allowing rapid and high-resolution documentation of historical locations [24–27]. Additionally, it enables the recording and tracking of antique and historical architectural structures that are unreachable due to endangered species or have been permanently destroyed. Satellite and aerial images are becoming increasingly abundant and diverse, making them ideal for this subject. However, these massive quantities of historical architectural information are still examined via human labeling and visual examination. Numerous studies have been conducted on the implementation of deep learning to various classifications of images, both generic and specialized in different areas [28–31]. Additionally, there are several projects devoted to the categorization of photographs of architectural heritage, trying to use different methods [32–35]. Since 2016, research utilizing CNNs to automate identifying historical architectural and archaeological elements has made significant progress in this field. In this study, the use of CNNs for prospecting vernacular architecture is demonstrated by extracting characteristics of central courtyards from aerial and satellite images of ancient cities.

### 1.3. State of Art

#### 1.3.1. Central Courtyards Studies

The primary goal of this research is to examine the traditional central courtyard idea as a passive cooling technique for enhancing interior thermal conditions in Iran's BS climate. An empirical survey was conducted to analyze three significant courtyard architecture variants, including their orientation, size, and ratios, as well as their opaque and transparent surfaces, in fourteen significant historic buildings in five ancient cities, including Mashhad, Shiraz, and Tehran in the BSks mesoclimate, as well as Dezful and Shushtar in the BSs-mesoclimate zone. The analysis considered three design variations: (1) the courtyard's orientation, extension, and rotation angle; (2) the courtyard's size and ratios concerning the overall wind flow; and (3) the courtyard's opaque and transparent surfaces concerning entry points to the overall wind flow [36]. The examination of criteria 1 for the Nasir al Molk building in Shiraz, one of the fourteen study cases, is depicted

in Figure 4. Both the house and courtyard were oriented northeast-southwest, with the courtyard angled 45 degrees to the north.



**Figure 4.** Examination of criterion 1 for the Nasir al Molk building in Shiraz, (a) courtyard alignment and rotation angle, (b) building alignment and rotation angle [36].

This research uses the study findings of fourteen courtyard buildings as case studies to determine the design patterns of courtyard homes at the city level. They summarize the findings of the criterion analysis to offer an acceptable model for building courtyard homes in these cities. Additionally, effort was made to suggest a design based on the length, width, and height ratios (Table 1).

**Table 1.** Summary of the findings from the three-criteria study conducted in Mashhad, Shiraz, Tehran, Dezful, and Shushtar. Furthermore, Equations for designing courtyards and the connections between their length and width at various sizes [36].

Zone	Equation	Mesoclimate	Equation	City	Equation
BS	$W = 0.8 L + 0.17$	BSks	$W = 0.73 L + 2.4$	Mashhad	$W = 0.7 L + 3.67$
				Shiraz	$W = -0.04 L - 13.9$
		BShs	$W = 0.85 L - 2.8$	Tehran	$W = 1.17L - 4$
				Dezful	$W = -0.1 L + 8.26$
				Shushtar	$W = 0.99 L - 5.49$

According to this study, approximately 17% of the courtyard should be dedicated to natural features, 8% water, and 9% plants. This ratio is feasible for ensuring adequate thermal comfort in the courtyard and adjacent regions based on observation and prior research. This quantitative research indicates that Iranian traditional central courtyards were built with an emphasis on orientation and geometry, concerning physical and natural factors to function well passive cooling systems. Soflaei et al. used the same technique in a prior study on various cities and climatic zones to evaluate the physical characteristics of six significant traditional courtyard homes in Iran's BWks-mesoclimate area. Findings indicated that most Iranian courtyards examined were purposefully built to allow direction, size, and ratio to function as microclimate modifiers.

### 1.3.2. Aerial Images Remote Sensing

This article compares several techniques for automatically detecting buildings in aerial photos and laser data with varying spatial resolutions. Five techniques are evaluated in two research areas utilizing features derived at the pixel and object level, with the precondition that all methods use the same training set. The techniques are evaluated using error measurements derived by superimposing the findings on every area's manually produced reference map [37]. Land cover classification, using very high-resolution remote sensing imagery, appears to progress from land-use categorization to pixel-level segmentation. Inspired by the current breakthrough of deep learning and the filtering technique used in computer vision, this study proposes a segmentation method that utilizes deep residual networks to create an image segmentation neural network and a guided filter to exclude

buildings from remote sensing photos [38]. This study develops a GRRNet for building extraction by fusing high-resolution aerial imagery with LiDAR point clouds. The updated residual learning network is utilized as an encoder element of GRRNet to train multi-level properties from the fusion data. A GFL unit is implemented to avoid inadvertently sending features and maximize the categorization effects [39].

The first stage of categorization, the extraction function, was performed in this study through a pre-trained CNN that recovers the deep characteristics of aerial pictures from different network layers, including the mean pooling layer and any prior convolutional layer. After dimensional reduction on large vector features, the concatenation algorithm was expanded to include the derived features from various neural networks. The research team experimented with different CNN designs to obtain the best results [40]. This article summarizes the results of a study on identifying qanat shafts using CORONA satellite imagery from the Cold War period. The application of deep CNNs for automated archaeological feature detection was investigated through remote sensing. Their case study involved the qanat structure in Iraq [41].

The efficient non-local residual U-shape network is a new type of network proposed in this research. It consists of a well-designed U-shape encoder/decoder setup and an optimized non-local block termed the asymmetric non-local block pyramid. The encoder/decoder framework is utilized to extract and reconstruct feature charts with great care, and APNB will gather global contextual information through a self-care method. The suggested ENRU-Net is evaluated and compared to existing state-of-the-art models using two widely utilized public aerial image building datasets [42]. This study suggested a novel deep transfer learning technique for automatically detecting Hakka Weilong Houses, a renowned ancient dwelling and a major cultural icon for Hakka, a Chinese minority found worldwide. The RS produced the Hakka Weilong House Image Dataset utilizing aerial photos of urban and suburban Meizhou [43].

The article described the development of a novel, fully automated three-dimensional building restoration system capable of producing first-level detail building models from multi-view aerial images without relying on additional information. The recovered models in this study come close to the accuracy and reliability of manually defined models. Three components comprise the proposed method: (1) efficient dense alignment and reconstruction of the earth's surface; (2) consistent building footprint recovery and polygon regularization; and (3) extremely accurate building rooftop and base height inference [44].

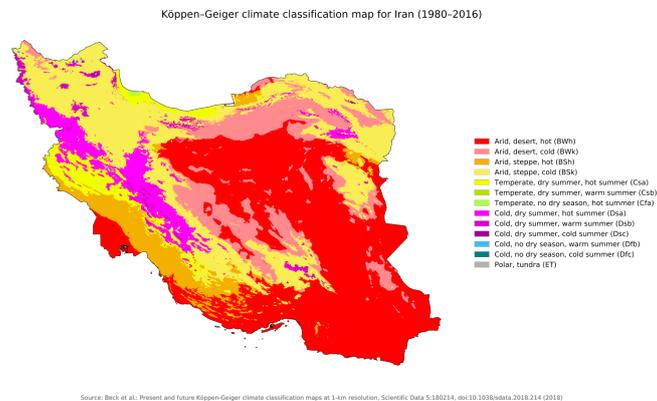
#### 1.4. Research Goals

Numerous research initiatives in Iran concentrate on ancient central courtyards. They compared numerous characteristics of central courtyards and homes, including their orientation and ratio, in a series of case studies [20,36,45–48]. However, conventional data collecting techniques restricted researchers from conducting a thorough case study of nearly entire central courtyards in a municipality or across several cities in Iran. This study will overcome this restriction via machine learning techniques to (VHR) aerial imagery. It will gather and analyze large amounts of data on central courtyard characteristics to uncover previously undiscovered meteorological and geometrical variables that influenced the design of central courtyards. Additionally, this research analyzes the last dataset for the Iranian historical central courtyard as a passive cooling technique in BS and BW climates, to offer an architectural model for modern sustainable structures. The finished design model may serve as a reference for designers and researchers as they progress in the historic districts of these Iranian cities.

#### 1.5. Case Study

The case study for this research involves nine cities in Iran's various climate zones (Figure 5 based on [49]). These cities have different climate conditions for gathering big data of central courtyards. Therefore, the samples will have a variety of features, according to the climate zone. Table 2 shows the list of the cities in the case study and the climate

divisions. These cities are separated into two main climate zones BS and BW: arid/steppe and arid/desert. BS and BW climates also have two subcategory climate zones: “H and k” in the cold and hot categories. Thus, there are four climate categories, called Bsh, Bsk, Bwh, and Bwk.



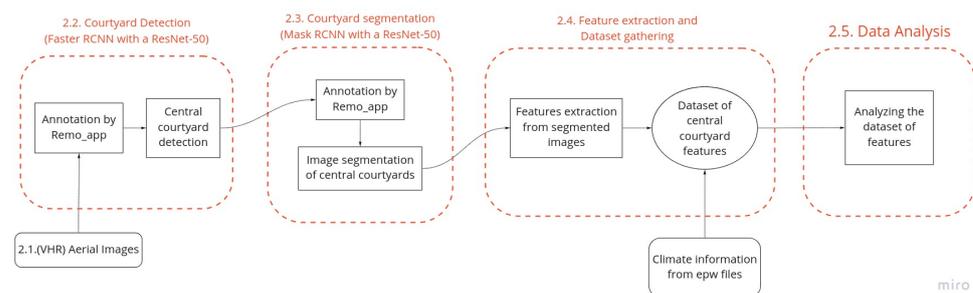
**Figure 5.** The Köppen-Geiger climate classification map for different areas of Iran [50].

**Table 2.** Climate classification of the nine cities in the case study located in various parts of Iran.

Climate Divisions	Bsh	Bsk	Bwh	Bwh
	Shiraz	Abarkooch Esfahan	Bushehr Semnan Kong	Birjand Meybod Yazd

## 2. Material and Method

Two distinct research methods were employed in this work. The first involved library research that concentrated on sustainable historical architecture and the passive cooling effect of central courtyards. The second research method was a survey that focused on the physical environmental characteristics of close to 30,000 traditional courtyard homes in nine historic Iranian towns. This research study used a variety of machine learning techniques. The typical machine learning framework is composed of the following components: data gathering, preprocessing, model construction, and model validation. Data collection and preparation result in generating an annotated dataset based on ground truth data or expert opinion. Post-processing of data is required to identify out-layers or inaccuracies. The overall surveying component of the research is divided into four distinct stages. The stages are as follows: (1) detection of courtyards; (2) segmentation of courtyards; (3) extraction of features; and (4) data analysis. In summary, the procedure begins with extremely high-resolution aerial images and concludes with a wide assortment of approximately 30,000 samples, including 76 geometric and climatic characteristics. Figure 6 illustrates the surveying subsection of the study’s flow chart.



**Figure 6.** Research workflow.

### 2.1. Very High-Resolution (VHR) Aerial Images

Detecting objects in VHR aerial pictures is a challenging task. The approaches proposed in the literature for solving the object detection problem in VHR images are classified into two broad categories: traditional approaches that rely on handcrafted features and deep learning-based approaches that employ a CNN as a feature extractor and achieve superior performance. Handcrafted features have limited presentation potential and lack the required precision [51]. Deep learning has exceptional performance in various areas, including image processing, due to the automated creation of features [23,52–55]. In numerous benchmarks, including PASCAL [56] and COCO [54,55,57,58], region-based CNNs outperformed conventional object detection techniques [54,55,57–59].

However, object recognition is more straightforward in these benchmarks than it is in VHR aerial images. Natural photographs depict significantly larger objects than aerial photographs. Additionally, the appearance of objects changes significantly in VHR images due to occlusion, shadow, lighting, resolution, and perspective fluctuation. As a result, object identification in VHR aerial pictures is more challenging than in natural photographs (Figure 7).



**Figure 7.** Comparisons of a natural pictures from (a) the COCO dataset and (b) VHR aerial images [60].

The primary source of input data for this study involves a collection of VHR aerial images. These images were taken between 2011 and 2016 by Iran’s National Cartographic Center (Supplementary Materials National Cartographic Center). Aerial photographs of these nine cities are available in a variety of scales, resolutions, and directions. As a result, a preprocessing procedure was used to divide these aerial images into smaller images. Furthermore, the scale and direction of each aerial image of these cities were measured manually. Except for Esfahan, the aerial images were divided into 60 subdivided images. Esfahan’s aerial images were subdivided into 30 subimages due to their resolutions being less than half of the main images. The images were subdivided because their resolutions ( $23,080 \times 15,080$  pixels) were too high for the training and object detection model. Table 3 shows the list of the cities, their pixel scales, and their northern angles.

**Table 3.** The scale of the pixels and the north direction of each city.

City	Scale of Pixels in Meters	Northern Angle in Degrees
Abarkooh	0.075	−4.76
Birjand	0.083	−1.35
Bushehr	0.099	+30.39
Esfahan	0.058	−31.40
Kong	0.045	+35.23
Meybod	0.074	−68.90
Semnan	0.051	+26.03
Shiraz	0.087	−36.45
Yazd	0.071	−34.60

## 2.2. Central Courtyard Detection

For the training dataset, 57 subdivided images from all cities were selected. Next, 1565 courtyards were annotated by Remo app (Supplementary Materials Remo.ai: Image Datasets Management) to prepare the training and evaluation task dataset. During pre-processing, 20% of the dataset was randomly separated for the validation task and 80% for training the model. The data augmentation process increased the training dataset. The authors selected the training dataset from all nine cities to ensure that the training task included various central courtyards with distinct characters. Following that, the object detection model was trained using the PyTorch toolkit.

The article proposes an end-to-end object identification model for VHR aerial pictures that generates high-level multi-scale semantic feature maps with high-quality data for multi-scale object identification tasks. Extensive tests were conducted utilizing the ResNet-50 backbone [61]. The proposed model is composed of four different components. The first subsection is the backbone network, consisting of ResNet 50's convolutional blocks. The second subsection is the bottom-up route, which uses the last layer of the backbone network's convolutional blocks. The third subsection is the suggested densely linked feature pyramid network, which is the top-down route. The last subsection is the predictor head, which is used to predict classes and bounding boxes.

Residual networks [61] are deep convolutional networks in which blocks of convolutional layers are bypassed through shortcut connections. The fundamental blocks, dubbed "bottleneck" blocks, adhere to two straightforward design principles: (1) when the output feature map is the same size as the input feature map, the layers have the same number of filters. (2) When the size of the output feature map is halved, the number of filters is doubled. Downsampling is accomplished directly via convolutional layers with a stride 2, and batch normalization occurs immediately after each convolution and before ReLU activation. When the sizes of the input and output are identical, the identity shortcut is utilized. When the dimensions become larger, the projection shortcut aligns them through  $1 \times 1$  convolutions. When shortcuts traverse two-dimensional feature maps in either case, they do so with a stride 2. The network concludes with a 1000 fully connected layer activated using SoftMax.

In the proposed approach, a FasterRCNN-ResNet-50 Backbone model pre-trained on COCO-train2017 is used for object identification. Despite the distinction between natural and aerial images, it is assumed that ResNet-50 trained extensively on ImageNet can still be used to improve the effectiveness of courtyard identification tasks, as gathering and annotating significant numbers of courtyard samples remains a significant challenge. The first 49 layers of the ResNet-50 Backbone were transferred, pre-trained on COCO train2017, and left frozen on the courtyard detection task, utilizing transfer learning methods [62]. These layers can be viewed as layers for learning feature extraction.

The evaluation dataset was used to determine the final result of the object detection, and the result indicated an accuracy of approximately 90% on the evaluation dataset (Figure 8). The courtyard detection and crop function output are two different datasets of the courtyards, and the house belonging to the courtyard is called the offset dataset. Therefore, the crop function produces an image of the detected area in the subdivided image and an offset image three times bigger in width and height. For example, the second detected courtyard in the subdivided image "1176-R9\_21.jpg" is titled "1176-R9\_21\_1.jpg" and measures  $190 \times 258$  pixels in size. Furthermore, the offset image for this detected courtyard is named "1176-R9\_21\_1\_o.jpg" and measures  $479 \times 682$  pixels in size. These two examples of the courtyard and courtyard\_offset datasets are demonstrated in Figure 9. Following that, two datasets, courtyards and courtyard\_offset are prepared for image segmentation.



**Figure 8.** Samples of the validation dataset after the courtyard detection task.



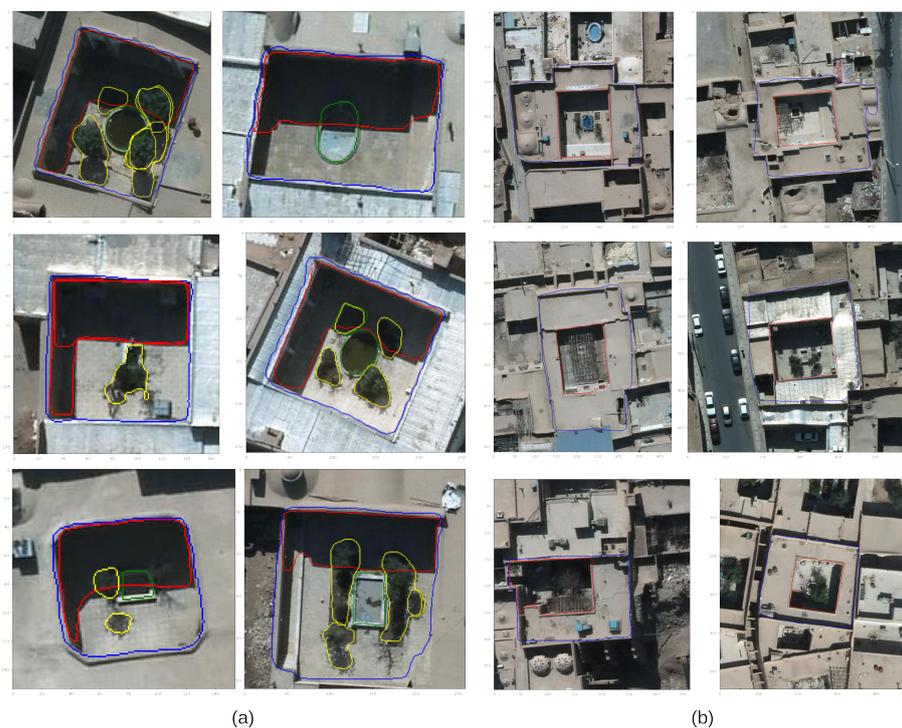
**Figure 9.** Examples of a sample with the courtyard\_ID: 1176-R9\_21\_1, in the courtyard dataset and the courtyard\_offset dataset.

### 2.3. Image Segmentation

The Remo app (Supplementary Materials Remo.ai: Image Datasets Management) was used in this study to perform the annotation task for the segmentation step. This step employs two distinct segmentation models: one for segmenting courtyards and the objects contained within, and another for segmenting offset images. Each training dataset contains 550 images; thus, the annotation task was completed for 1100 images in this step. For the first dataset, 4 objects were segmented for training the model: courtyard, a water place, green area, and shadow. Moreover, the courtyard and its associated house were segmented



a feature extractor often employed in two-stage object recognition [61,67]. For semantic segmentation, the features of each ROI are fed into basic convolutional layers and object masks. This study applies transfer learning to a mask R-CNN model from Detectron2 [68]. To this end, a ResNet-50 backbone pre-trained on COCO-train2017 is used [61]. After training the model with the two courtyards and courtyard\_offset datasets, the model was evaluated with a 20% test dataset. The evaluation demonstrates that courtyards, shadows, and houses achieve greater than 90% accuracy in classification, bounding-box, and masking tasks. However, the accuracy of water places and green areas was approximately 80%. Figure 12 illustrates several examples of evaluation datasets following the segmentation task. These images are segmented by the machine for testing the result of the training task. For example, the machine wrongly segmented the shadow of the green areas (yellow lines) in samples that show low accuracy, in some cases.



**Figure 12.** Examples of segmentation task evaluations on the testing dataset. (a) Segmentation of the courtyard (blue), shadow (red), green areas (yellow), and water places (green) in the courtyards dataset. (b) Segmentation of courtyard (red) and house (blue) in the courtyards\_offset datasets.

### 3. Results

#### 3.1. Feature Extraction and Dataset Gathering

The image segmentation was applied to the courtyards, and courtyard\_offset datasets, and the result was a set of polygons for each courtyard's objects. The next step was to extract features and collect datasets to create around 30,000 samples with 76 features to create the final comprehensive datasets. In this study, all polygons were combined from the same object class into a single object. Following that, each object's area in square meters was measured. The minimum bounding box for courtyards, houses, and water places was then identified for further measurements (Supplementary Materials MinimumBoundingBox). Figure 13 depicts a fitting minimum bounding box on a polygon. In addition, to extract the geometrical features of the selected object, the dimensions and direction of the minimum bounding box were measured.



**Figure 13.** An example of a polygon’s minimum bounding box. The red line represents the polygon created by the segmentation task that depicts the courtyard object. Furthermore, blue is the courtyard’s minimum bounding box.

Rather than geometric features, the dataset now includes several climatic features for each city. The EnergyPlus Weather Format (EPW) (Supplementary Materials Weather Data | EnergyPlus) files were then used to extract climatic data for each city. The EPW files contained the climatic information of each city for 14 years (2004–2018). After exporting the climatic data, the mean, min, max, and 50% were calculated for the 14 years and then added to the main dataset. As a result, the final features for each courtyard were 76 geometric and climatic data points. Table 4 summarizes these features in terms of their geometric and climatic classifications. The final dataset consisted of 29,191 samples collected from nine cities. Following that, the data were ready for additional analysis.

**Table 4.** The list of extracted features from the segmentation task (geometric) and EPW files (climatic).

Geometric (22 Features)	Climatic (54 Features)
Courtyard_area (m <sup>2</sup> ),	climate_divisions, latitude, longitude,
Water_place_area (m <sup>2</sup> ),	altitude, temp_air_mean,
Green_area (m <sup>2</sup> ),	temp_air_min, temp_air_50%,
Courtyard_length (m),	temp_air_max, temp_dew_mean,
Courtyard_width (m),	temp_dew_min, temp_dew_50%,
Courtyard_direction (deg),	temp_dew_max,
Water_place_length (m),	relative_humidity_mean,
Water_place_width (m),	relative_humidity_min,
Water_place_direction (deg),	relative_humidity_50%,
House_area (m <sup>2</sup> ),	relative_humidity_max,
House_length (m),	atmospheric_pressure_mean,
House_width (m),	atmospheric_pressure_min,
House_direction (deg),	atmospheric_pressure_50%,
City,	atmospheric_pressure_max, etr_mean,
Water_place/Courtyard_area_Ratio,	etr_min, etr_50%, etr_max,
Green/Courtyard_area_Ratio,	ghi_infrared_mean, ghi_infrared_min,
Courtyard/House_area_Ratio,	ghi_infrared_50%, ghi_infrared_max,
Water_place/House_area_Ratio,	dni_mean, dni_min, dni_50%, dni_max,
Green/House_area_Ratio,	dhi_mean, dhi_min, dhi_50%, dhi_max,
Courtyard_Ratio,	global_hor_illum_mean,
House_Ratio,	global_hor_illum_min,
Water_place_Ratio	global_hor_illum_50%,
	global_hor_illum_max,
	diffuse_horizontal_illum_mean,
	diffuse_horizontal_illum_min,
	diffuse_horizontal_illum_50%,
	diffuse_horizontal_illum_max,
	wind_direction_mean,
	wind_direction_50%,
	wind_speed_mean, wind_speed_min,
	wind_speed_50%, wind_speed_max,
	precipitable_water_mean,
	precipitable_water_min,
	precipitable_water_50%,
	precipitable_water_max,

### 3.2. Data Analysis

The first step of data analysis was data preprocessing. During data preprocessing, the out layers or incorrectly detected features were identified. For instance, there were several courtyards with an area of less than one (m<sup>2</sup>) in size; as a result, they were identified, and the entire sample from the dataset was removed. Numerous samples in the dataset contained incorrect data due to the impossibility of checking the results of courtyard detection and image segmentation after each step. However, it was straightforward at identifying and deleting this information from the final dataset. It was expected that approximately 20% of the dataset would contain incorrect data due to a 10% error in object detection and a 10% error during the image segmentation steps. Throughout the data analysis stage, an attempt was made to uncover relations between the dataset features by the use of the Pearson correlation coefficient.

### 3.3. Dataset Analysis of the Cities

The final dataset contained 29,190 samples and 77 columns, one of which was a courtyard ID column. The dataset was reduced to 26,437 samples after preprocessing and removing incorrect and outlier data. As a result, approximately 9.5% of the courtyard detection and segmentation were incorrect. Consequently, these records were deleted from their dataset. After data preprocessing, the final dataset contained 26,437 rows × 77 columns. The first step in reading a dataset is to visualize all of its values in a single plot to understand their distribution. As a result, using a heatmap across the entire dataset is a viable solution (Figure 14). It shows that, for example, the correlation between courtyard\_length and courtyard\_width is approximately 0.88, showing a strong relationship between these two variables. Furthermore, there is no correlation between the house\_ratio and latitude features, as the heatmap indicates that these features have a correlation coefficient of 0.033.

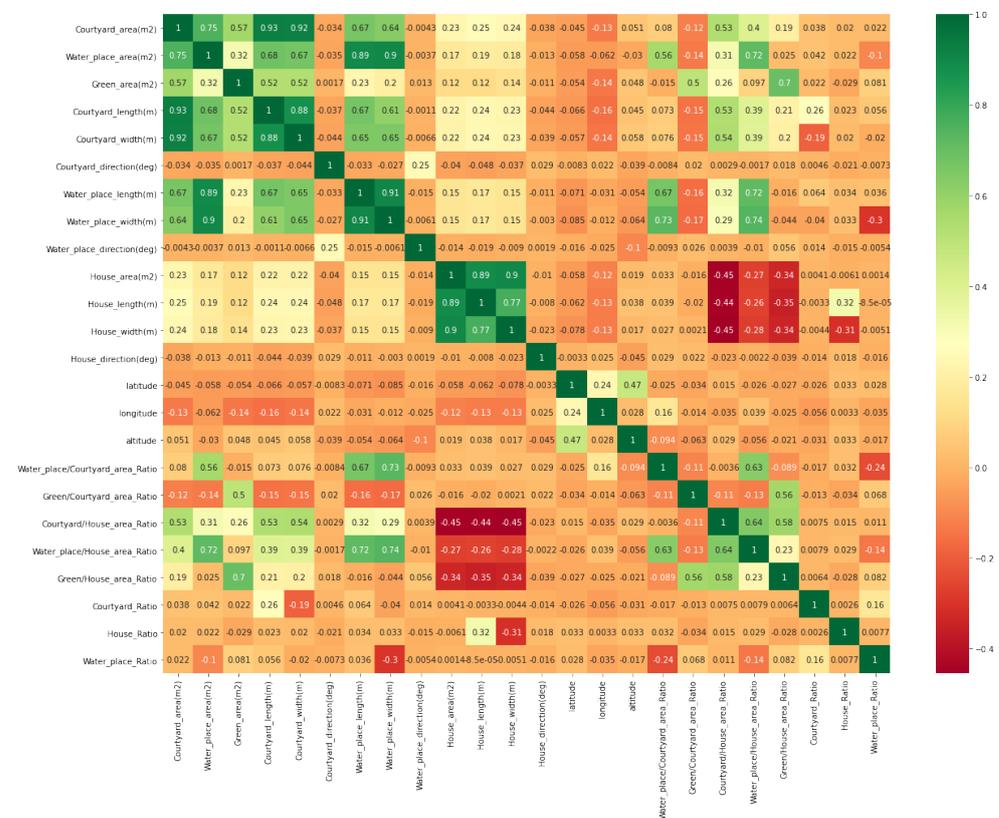
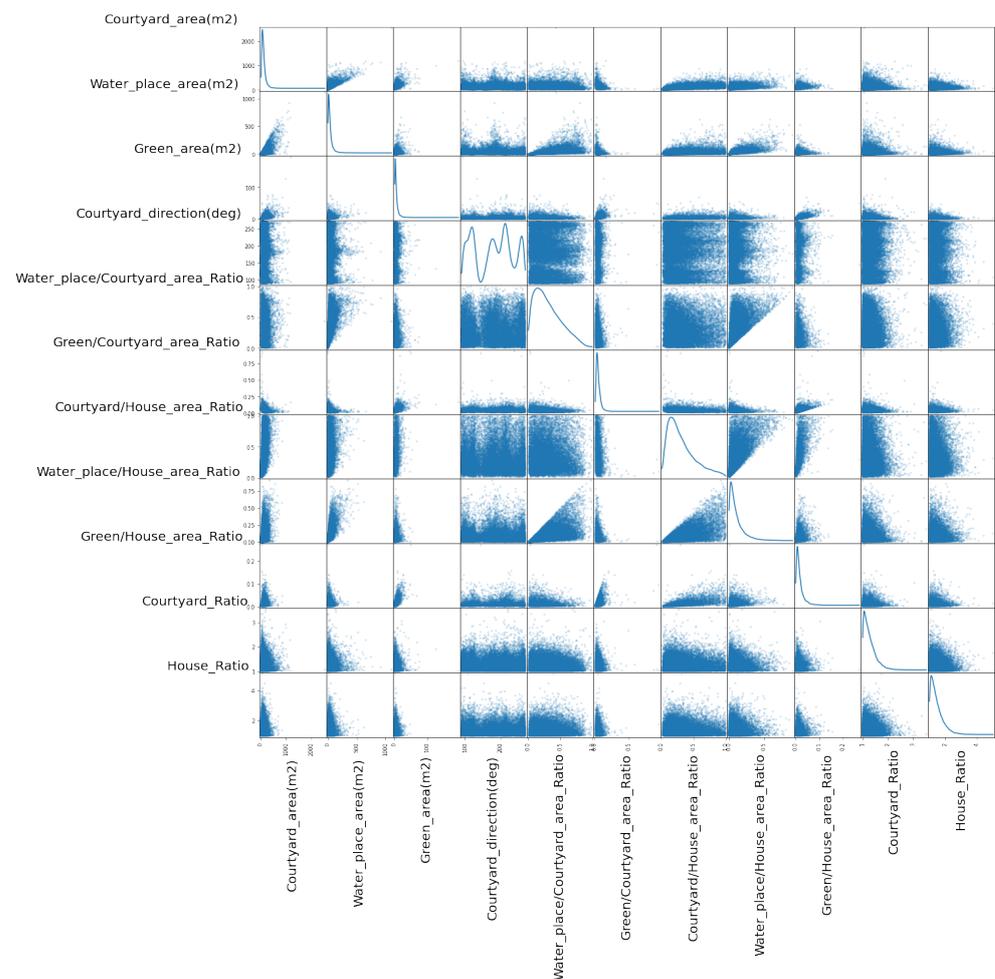


Figure 14. The heatmap of the dataset’s geometric features.

Another visualization method used in this research was scatter plots, which show how the data are affected by different features. For instance, the scatter plot of courtyard\_area and water\_place area demonstrates a logical spreading of data (Figure 15).

Within the visualization, different extracted features can be compared and analyzed in order to reach meaningful results. In this respect, the above visualizations are some examples of the implementation of the generated dataset. It should be mentioned that while reaching meaningful relationships among features is highly significant, and it is not the goal of this paper. Therefore, the mentioned graphs show the possibilities of using the datasets as an example. In this respect, the most important example of this dataset's outcome would be a model for the courtyard shape, which will be explained in the following section.



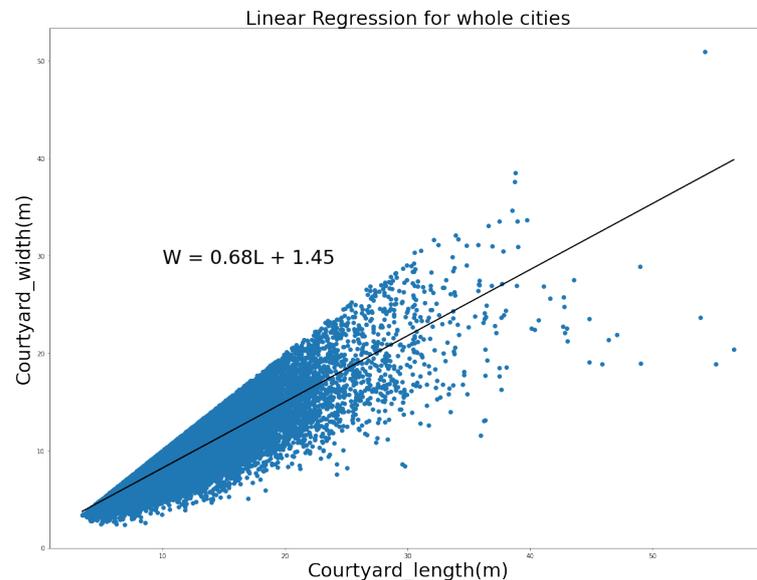
**Figure 15.** The scatter matrix of several geometrical features of the dataset.

#### 4. Discussion

##### *Proposed Model for the Dimension of the Central Courtyard*

One of the study's outcomes is a proposal for a design model for central courtyards, as a passive energy solution in these particular cities. Additionally, this dataset can pose as a model for additional cities, by running the central courtyard feature extraction model on other aerial images of various cities. The design model's purpose is to determine the optimal direction and dimension of a central courtyard in these cities or to forecast the shape of a destroyed central courtyard in the historical region of the cities represented in the datasets. It can also serve as a model for developing and reconstructing historical districts in heritage cities. For instance, in the historical region of Yazd, Iran, there is an empty field. The proposed design model in this study will predict the courtyard's optimal

direction, dimension, and area, based on thousands of samples taken throughout Yazd. It predicts these features through the use of data science prediction models, such as linear regression (Figure 16). The results reveal a strong correlation between the length and width of central courtyards in all cities, indicating a distinct pattern in the design of central courtyard houses throughout Iran's history.

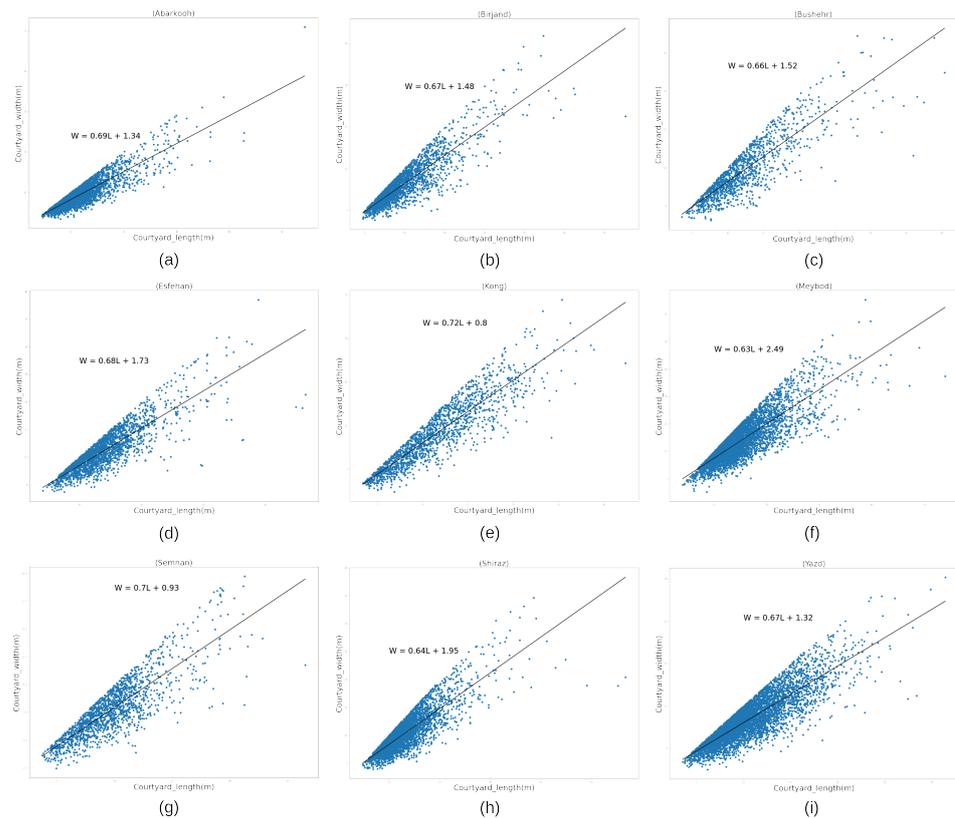


**Figure 16.** Relationship between length and width of the studied courtyards and the linear regression prediction equation.

Rather than a single mathematical equation, the dataset contains prediction equations for each climatic zone and city. It shows how the proportion of central courtyards varies by city and climatic zone (Table 5 and Figure 17). The mathematical equation for linear regression varies in weight between 0.63 and 0.72 in various cities. Additionally, the second column of this table included an equation for each of the climatic zones. Furthermore, the final mathematical equation for the dataset's entire courtyards is included in the first column of this table.

**Table 5.** The mathematical equations for calculating the length and width of courtyards in various cities and climatic zones.

Dataset (Figure 16)		Climate Zones		Cities (Figure 17)	
Dataset	$W = 0.68L + 1.45$	BSh	(a) $W = 0.64L + 1.95$	Shiraz	(H) $W = 0.64L + 1.95$
		Bsk	(b) $W = 0.69L + 1.45$	Abarkooh	(a) $W = 0.69L + 1.34$
				Esfahan	(d) $W = 0.68L + 1.73$
		BWh	(c) $W = 0.7L + 1.04$	Bushehr	(c) $W = 0.66L + 1.52$
				Semnan	(g) $W = 0.7L + 0.93$
				Kong	(e) $W = 0.72L + 0.8$
		BWk	(d) $W = 0.67L + 1.53$	Meybod	(f) $W = 0.63L + 2.49$
				Birjand	(b) $W = 0.67L + 1.48$
				Yazd	(i) $W = 0.67L + 1.32$



**Figure 17.** The scatter plot and linear regression line depict the length and width of courtyards in various cities included in the dataset.

## 5. Conclusions

The features of central courtyards were extracted in this study, in order to collect big data for future research. The flow chart in Figure 6 indicates that this study is divided into four distinct steps. These steps are as follows: (1) preprocessing VHR aerial images for object detection and detection of courtyards using a faster RCNN model; (2) segmentation of the objects in the courtyards using mask RCNN; (3) extraction of features associated with each object and collection of big data; and (4) data analysis and data science of the dataset to discover mathematical relations between features. In summary, the procedure begins with extremely high-resolution aerial images and concludes with a vast collection of 26,437 samples, each of which contains 76 geometric and climatic features. Additionally, the final data analysis models revealed several relations and correlations between the final dataset features. It demonstrated that, while the direction and dimension ratios of central courtyards vary across cities and climatic zones, they all serve a purpose. Furthermore, the proposed design model can guide researchers and designers working in historic cities. The model can predict and identify the optimal form of a central courtyard in a historic region based on thousands of courtyard samples. In summary, this paper postulates that machine learning techniques used in computer vision and data science are highly beneficial for studying historic architecture. This research demonstrates that it is possible to study over 26,000 courtyard samples in a dataset using only a single computational method.

In this research, the authors faced several technical and practical challenges, and in order to minimize the challenging areas and consider the available computational resources, some limitations were considered. Since detecting water places and green areas needed further tuning, and this task was not among the main goals of this research, the authors did not focus on this part. Apart from that, courtyards located between two subdivided images were not considered in the dataset as it reduced the model's accuracy. Finally, the VHR aerial images do not cover all historical cities in Iran or the Middle East, which is why a limited number of cities could be selected as the case study.

For future studies, extending the dataset over the borders of Iran would be recommended to study the impact of this style of architecture on other Middle Eastern countries. Moreover, detecting more details in the courtyards, such as their height, could add value to this dataset. In terms of analyzing the dataset, focusing on each dual relationship of features can reveal details about the principles of the central courtyard's design in this style of architecture.

**Supplementary Materials:** The following are available at <https://www.mdpi.com/article/10.3390/rs13234843/s1>. National Cartographic Center; Remo.ai: Image Datasets Management; Minimum-BoundingBox; Weather Data | EnergyPlus.

**Author Contributions:** Conceptualization, H.Y.; methodology, H.Y.; software, H.Y.; validation, H.Y.; formal analysis, H.Y.; investigation, H.Y.; resources, H.Y., I.V., M.B. and S.M.; data curation, H.Y.; writing—original draft preparation, H.Y.; writing—review and editing, H.Y., I.V., A.N., R.S. and H.B.; visualization, H.Y.; supervision, I.V., A.N., M.B., S.M., R.S. and H.B.; funding acquisition, I.V., M.B., A.N. and H.B. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Poznan University of Technology, within the framework of the research project entitled “Mapping of architectural space, the history, theory, practice, contemporaneity”, grant number 0112/SBAD/0185. This research was founded by the EUNICE (European University for Customized Education) project; Poznan University of Technology (Poland), and Brandenburg University of Technology Cottbus-Senftenberg (Germany).

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

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The dataset generated in this paper is accessible via <https://github.com/hadi-yazdi/yazd>, accessed on 5 October 2021. The object detection and segmentation training codes consist of the final trained models for courtyard detection and segmentation, the final dataset, data analysis, and model fitting codes.

**Acknowledgments:** This paper is a part of the first author's thesis at the Poznan University of Technology and Brandenburg University of Technology. The authors also express their gratitude to Wendland, for his support and assistance with this research.

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

## References

1. Bonenberg, W. The Role of Cultural Heritage in Sustainable Development. Values and Valuation as Key Factors in Spatial Planning of Rural Areas. In *Advances in Human Factors in Architecture, Sustainable Urban Planning and Infrastructure*; Charytonowicz, J., Falco, C., Eds.; Springer International Publishing: Washington, DC, USA, 2020; pp. 124–134.
2. Bazazzadeh, H.; Nadolny, A.; Attarian, K.; Safar ali najar, B.; Hashemi safaei, S. Promoting sustainable development of cultural assets by improving users' perception through space configuration; Case study: The industrial heritage site *Sustainability* **2020**, *12*, 5109. [\[CrossRef\]](#)
3. Mahdavinejad, M.; Didehban, M.; Bazazzadeh H. Contemporary architectural heritage and industrial identity in historic districts, case study: Dezful. *J. Stud.-Iran.-Islam. City* **2016**, *6*, 41–50.
4. Pazder, D.; Kaźmierczak, B. Responsibility in the revitalization process—The aspect of the economic policy. *Sci. Pap. Poznań Univ. Technol. Archit. Urban Plan.* **2009**, *17*, 207–210.
5. Pazder, D. The revitalization of the downtown cultural space as the factor of increasing the city attractiveness. *Sci. Pap. Poznań Univ. Technol. Archit. Urban Plan.* **2010**, *22*, 123–127.
6. Sahebzadeh, S.; Heidari, A.; Kamelnia, H.; Baghbani, A. Sustainability Features of Iran's Vernacular Architecture: A Comparative Study between the Architecture of Hot–Arid and Hot–Arid–Windy Regions. *Sustainability* **2017**, *9*, 749. [\[CrossRef\]](#)
7. Bazazzadeh, H.; Mehan, A.; Nadolny, A.; Hashemi Safaei, S. The Importance of Flexibility in Adaptive Reuse of Industrial Heritage: Learning from Iranian Cases. *Int. J. Conserv. Sci.* **2021**, *12*, 113–128.
8. Foruzanmehr, A. People's perception of the loggia: A vernacular passive cooling system in Iranian architecture. *Sustain. Cities Soc.* **2015**, *19*, 61–67. [\[CrossRef\]](#)
9. Bazazzadeh, H. Truth of sincerity and authenticity or lie of reconstruction; Whom do the visitors of cultural heritage trust? In *Defining the Architectural Space—The Truth and Lie of Architecture*; Battaino, C., Bonenberg, A., Dal Fabbro, A., Juzwa, N., Kobylarczyk, J., Malacarne, G., Segal, R., Styk, J., Eds.; ATUT: Kraków, Poland, 2019; Volume 9, pp. 7–18.
10. Ragette, F. *Traditional Domestic Architecture of the Arab Region*; Google-Books-ID: dLWm6eVwnwAC; Edition Axel Menges: Zlín, Czech Republic, 2003; ISBN 978-3-932565-30-4.

11. Soleymanpour, R.; Parsaee, N.; Banaei, M. Climate Comfort Comparison of Vernacular and Contemporary Houses of Iran. *Procedia—Soc. Behav. Sci.* **2015**, *201*, 49–61. [[CrossRef](#)]
12. Bazazzadeh, H.; Nadolny, A.; Hashemi Ssafaei, S. Climate Change and Building Energy Consumption: A Review of the Impact of Weather Parameters Influenced by Climate Change on Household Heating and Cooling Demands of Buildings. *Eur. J. Sustain. Dev.* **2021**, *10*, 1–12. [[CrossRef](#)]
13. Bazazzadeh, H.; Pilechiha, P.; Nadolny, A.; Mahdaveinejad, M.; Hashemi Safaei, S. The Impact Assessment of Climate Change on Building Energy Consumption in Poland. *Energies* **2021**, *14*, 4084. [[CrossRef](#)]
14. Tabrizikahou, A.; Nowotarski, P. Mitigating the Energy Consumption and the Carbon Emission in the Building Structures by Optimization of the Construction Processes. *Energies* **2021**, *14*, 3287. [[CrossRef](#)]
15. Khalili, M.; Amineldar, S. Traditional solutions in low energy buildings of hot-arid regions of Iran. *Sustain. Cities Soc.* **2014**, *13*, 171–181. [[CrossRef](#)]
16. Shahamat, H. Formal Sustainability in Traditional Architecture of Iran According to Five Principles of Traditional Architecture of Iran. *J. Appl. Environ. Biol. Sci.* **2014**, *4*, 100–110.
17. Abouei, R. Conservation of Badgirs and Qanats in Yazd, Central Iran. In Proceedings of the 23rd Conference on Passive and Low Energy Architecture, Geneva, Switzerland, 6–8 September 2006.
18. Tavassoli, M. *Urban Structure and Architecture in the Hot Arid Zone of Iran*; University of Tehran Press: Tehran, Iran, 1982.
19. Keshtkaran, P. Harmonization Between Climate and Architecture in Vernacular Heritage: A Case Study in Yazd, Iran. *Procedia Eng.* **2011**, *21*, 428–438. [[CrossRef](#)]
20. Amiriparyan, P.; Kiani, Z. Analyzing the Homogenous Nature of Central Courtyard Structure in Formation of Iranian Traditional Houses. *Procedia—Soc. Behav. Sci.* **2016**, *216*, 905–915. [[CrossRef](#)]
21. Memariyan, G. *Introduction to House Typology in Iran, House without Courtyard*; University of Science and Technology: Tehran, Iran, 1990.
22. Mahdaveinejad, M.; Yazdi, H. Daylightophil Approach towards High-Performance Architecture for Hybrid-Optimization of Visual Comfort and Daylight Factor in BSk. *Int. J. Archit. Environ. Eng.* **2017**, *11*, 1324–1327.
23. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*, 1097–1105. [[CrossRef](#)]
24. Lasaponara, R.; Masini, N. (Eds.) *Satellite Remote Sensing: A New Tool for Archaeology*; Remote Sensing and Digital Image Processing; Springer: Dordrecht, The Netherlands, 2012. [[CrossRef](#)]
25. Leisz, S.J. An Overview of the Application of Remote Sensing to Archaeology During the Twentieth Century. In *Mapping Archaeological Landscapes from Space*; Comer, D.C., Harrower, M.J., Eds.; SpringerBriefs in Archaeology; Springer: New York, NY, USA, 2013; pp. 11–19. [[CrossRef](#)]
26. Parcak, S.H. *Satellite Remote Sensing for Archaeology*; Routledge: Abingdon-on-Thames, UK, 2009. [[CrossRef](#)]
27. Wiseman, J.R.; El-Baz, F. (Eds.) *Remote Sensing in Archaeology*; Interdisciplinary Contributions to Archaeology; Springer: New York, NY, USA, 2007. [[CrossRef](#)]
28. Li, C.; Min, X.; Sun, S.; Lin, W.; Tang, Z. DeepGait: A Learning Deep Convolutional Representation for View-Invariant Gait Recognition Using Joint Bayesian. *Appl. Sci.* **2017**, *7*, 210. [[CrossRef](#)]
29. Tajbakhsh, N.; Shin, J.Y.; Gurudu, S.R.; Hurst, R.T.; Kendall, C.B.; Gotway, M.B.; Liang, J. Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning? *IEEE Trans. Med. Imaging* **2016**, *35*, 1299–1312. [[CrossRef](#)]
30. Pedraza, A.; Bueno, G.; Deniz, O.; Cristóbal, G.; Blanco, S.; Borrego-Ramos, M. Automated Diatom Classification (Part B): A Deep Learning Approach. *Appl. Sci.* **2017**, *7*, 460. [[CrossRef](#)]
31. Sa, I.; Ge, Z.; Dayoub, F.; Upcroft, B.; Perez, T.; McCool, C. DeepFruits: A Fruit Detection System Using Deep Neural Networks. *Sensors* **2016**, *16*, 1222. [[CrossRef](#)]
32. Goel, A.; Juneja, M.; Jawahar, C.V. Are buildings only instances? exploration in architectural style categories. In *ICVGIP '12: Proceedings of the Eighth Indian Conference on Computer Vision, Graphics and Image Processing, Mumbai, India, 16–19 December 2012*; Association for Computing Machinery: New York, NY, USA, 2012; pp. 1–8. [[CrossRef](#)]
33. Zhang, L.; Song, M.; Liu, X.; Sun, L.; Chen, C.; Bu, J. Recognizing architecture styles by hierarchical sparse coding of blocklets. *Inf. Sci.* **2014**, *254*, 141–154. [[CrossRef](#)]
34. Chu, W.T.; Tsai, M.H. Visual pattern discovery for architecture image classification and product image search. In *ICMR '12: Proceedings of the 2nd ACM International Conference on Multimedia Retrieval, Hong Kong, China, 5–8 June 2012*; Association for Computing Machinery: New York, NY, USA, 2012; pp. 1–8. [[CrossRef](#)]
35. Oses, N.; Dornaika, F.; Moujahid, A. Image-Based Delineation and Classification of Built Heritage Masonry. *Remote Sens.* **2014**, *6*, 1863–1889. [[CrossRef](#)]
36. Soflaei, F.; Shokouhian, M.; Mofidi Shemirani, S.M. Investigation of Iranian traditional courtyard as passive cooling strategy (a field study on BS climate). *Int. J. Sustain. Built Environ.* **2016**, *5*, 99–113. [[CrossRef](#)]
37. Khoshelham, K.; Nardinocchi, C.; Frontoni, E.; Mancini, A.; Zingaretti, P. Performance evaluation of automated approaches to building detection in multi-source aerial data. *ISPRS J. Photogramm. Remote Sens.* **2010**, *65*, 123–133. [[CrossRef](#)]
38. Xu, Y.; Wu, L.; Xie, Z.; Chen, Z. Building Extraction in Very High Resolution Remote Sensing Imagery Using Deep Learning and Guided Filters. *Remote Sens.* **2018**, *10*, 144. [[CrossRef](#)]

39. Huang, J.; Zhang, X.; Xin, Q.; Sun, Y.; Zhang, P. Automatic building extraction from high-resolution aerial images and LiDAR data using gated residual refinement network. *ISPRS J. Photogramm. Remote Sens.* **2019**, *151*, 91–105. [[CrossRef](#)]
40. Petrovska, B.; Zdravevski, E.; Lameski, P.; Corizzo, R.; Štajduhar, I.; Lerga, J. Deep Learning for Feature Extraction in Remote Sensing: A Case-Study of Aerial Scene Classification. *Sensors* **2020**, *20*, 3906. [[CrossRef](#)] [[PubMed](#)]
41. Soroush, M.; Mehrtash, A.; Khazraee, E.; Ur, J.A. Deep Learning in Archaeological Remote Sensing: Automated Qanat Detection in the Kurdistan Region of Iraq. *Remote Sens.* **2020**, *12*, 500. [[CrossRef](#)]
42. Wang, S.; Hou, X.; Zhao, X. Automatic Building Extraction From High-Resolution Aerial Imagery via Fully Convolutional Encoder-Decoder Network With Non-Local Block. *IEEE Access* **2020**, *8*, 7313–7322. [[CrossRef](#)]
43. Xiong, Y.; Chen, Q.; Zhu, M.; Zhang, Y.; Huang, K. Accurate Detection of Historical Buildings Using Aerial Photographs and Deep Transfer Learning. In Proceedings of the IGARSS 2020—2020 IEEE International Geoscience and Remote Sensing Symposium, Waikoloa, HI, USA, 26 September–2 October 2020; pp. 1592–1595. [[CrossRef](#)]
44. Yu, D.; Ji, S.; Liu, J.; Wei, S. Automatic 3D building reconstruction from multi-view aerial images with deep learning. *ISPRS J. Photogramm. Remote Sens.* **2021**, *171*, 155–170. [[CrossRef](#)]
45. Memarian, G.; Brown, F.E. Climate, culture, and religion: Aspects of the traditional courtyard house in Iran. *J. Archit. Plan. Res.* **2003**, *20*, 181–198.
46. Soflaei, F.; Shokouhian, M.; Mofidi Shemirani, S.M. Traditional Iranian courtyards as microclimate modifiers by considering orientation, dimensions, and proportions. *Front. Archit. Res.* **2016**, *5*, 225–238. [[CrossRef](#)]
47. Zarei, E.M.; Ashkezari, S.F.M.; Yari, M. The investigation of the function of the central courtyard in moderating the harsh environmental conditions of a hot and dry climate (Case study: City of Yazd, Iran). *Spatium* **2018**, *2017*, 1–9. [[CrossRef](#)]
48. Hajian, M.; Alitajer, S.; Mahdavinejad, M. The Influence of Courtyard on the Formation of Iranian Traditional Houses Configuration in Kashan. *Arman. Archit. Urban Dev.* **2020**, *13*, 43–55. [[CrossRef](#)]
49. Razinei, T. Köppen-Geiger climate classification of Iran and investigation of its changes during 20th century. *J. Earth Space Phys.* **2017**, *43*, 419–439. [[CrossRef](#)]
50. Razinei, T. Köppen-Geiger Climate Classification Map for Iran. 2018. Available online: [https://www.researchgate.net/figure/Koeppen-Geiger-climate-classification-of-Iran-22\\_fig1\\_341741246](https://www.researchgate.net/figure/Koeppen-Geiger-climate-classification-of-Iran-22_fig1_341741246) (accessed on 4 October 2021).
51. Dai, D.; Yang, W. Satellite Image Classification via Two-Layer Sparse Coding With Biased Image Representation. *IEEE Geosci. Remote Sens. Lett.* **2011**, *8*, 173–176. [[CrossRef](#)]
52. Zhang, D.; Meng, D.; Han, J. Co-Saliency Detection via a Self-Paced Multiple-Instance Learning Framework. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 865–878. [[CrossRef](#)]
53. Tian, Y.; Chen, C.; Shah, M. Cross-View Image Matching for Geo-Localization in Urban Environments. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 3608–3616.
54. Girshick, R. Fast R-CNN. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1440–1448.
55. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [[CrossRef](#)] [[PubMed](#)]
56. Everingham, M.; Eslami, S.M.A.; Van Gool, L.; Williams, C.K.I.; Winn, J.; Zisserman, A. The Pascal Visual Object Classes Challenge: A Retrospective. *Int. J. Comput. Vis.* **2015**, *111*, 98–136. [[CrossRef](#)]
57. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. SSD: Single Shot MultiBox Detector. In *Computer Vision—ECCV 2016*; Lecture Notes in Computer Science; Leibe, B., Matas, J., Sebe, N., Welling, M., Eds.; Springer International Publishing: Cham, Switzerland, 2016; pp. 21–37. [[CrossRef](#)]
58. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
59. Lin, T.Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft COCO: Common Objects in Context. In *Computer Vision—ECCV 2014*; Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2014; pp. 740–755. [[CrossRef](#)]
60. Tayara, H.; Chong, K.T. Object Detection in Very High-Resolution Aerial Images Using One-Stage Densely Connected Feature Pyramid Network. *Sensors* **2018**, *18*, 3341. [[CrossRef](#)]
61. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [[CrossRef](#)]
62. Yosinski, J.; Clune, J.; Bengio, Y.; Lipson, H. How transferable are features in deep neural networks? *arXiv* **2014**, arXiv:1411.1792.
63. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask R-CNN. *arXiv* **2018**, arXiv:1703.06870v3.
64. Jeong, D.; Kim, Y. Keypoint-based Deep Learning Approach for Building Footprint Extraction Using Aerial Images. *Korean J. Remote Sens.* **2021**, *37*, 111–122. [[CrossRef](#)]
65. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask R-CNN. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2980–2988. [[CrossRef](#)]
66. Zhang, X. Simple Understanding of Mask RCNN. 2018. Available online: <https://alittlepain833.medium.com/simple-understanding-of-mask-rcnn-134b5b330e95> (accessed on 5 October 2021).

- 
67. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *Adv. Neural Inf. Process. Syst.* **2015**, *28*, 91–99. [[CrossRef](#)] [[PubMed](#)]
  68. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Region-Based Convolutional Networks for Accurate Object Detection and Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *38*, 142–158. [[CrossRef](#)] [[PubMed](#)]