ISPRS Journal of Photogrammetry and Remote Sensing 175 (2021) 298–310

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Detecting individual abandoned houses from google street view: A hierarchical deep learning approach

Shengyuan Zou, Le Wang *

Department of Geography, University at Buffalo, the State University of New York, Amherst, NY 14261, United States

ARTICLE INFO

Keywords:
Residential housing abandonment
Street view
Knowledge-guided deep learning
Patch-based classification

ABSTRACT

Abandoned houses (AH) are focal points in urban communities by threatening local security, destroying housing markets, and burdening government finance in the U.S. legacy cities. In particular, individual-level AH detection provides essential information for fine-resolution urban studies, government decision-makers, and private sector practitioners. However, three primary conventional data sources (field data, utility data, and remote sensing data) cannot suffice to collect such fine-resolution data in the large spatial area via a cost-effective approach. To this end, Google Street View (GSV) imagery, which emerges as the mainstream open-access data source with global coverage, provides an opportunity to address this issue. Subsequently, a follow-up challenge confronting the detection of AH arises from the fact that it lacks an effective method that can discern authentic visual features from the redundant noise in GSV images. In this study, we aim to develop an effective method to detect individual-level AH from GSV imagery. Specifically, we developed a new hierarchical deep learning method to leverage both global and local visual features of AH in the detection. The method can be further divided into three steps: (1) Scene-based classification that can extract global visual features of AH was implemented through fine-tuning a pre-trained deep convolutional neural network (CNN) model. (2) We developed a patch-based classification method that can extract specific local features of AH. In this method, patches were generated from GSV images based on auto-detected local features, followed by being labeled as three categories: building patches, vegetation patches, and others. Two deep CNN models were employed to identify deteriorated building façade patches and overgrown vegetation patches, respectively. (3) Individual-level AH were detected by integrating scene classification results and patch classification results in a decision-tree model. Experimental results showed that the F-score of AH was 0.84 in a well-prepared dataset collected from five different Rust Belt cities. The proposed hierarchical deep learning approach effectively improved the accuracy comparing with the traditional scene-based method. In addition, the proposed method was applied to generate an AH map in a new site in Detroit, MI. Our study demonstrated the feasibility of GSV imagery in AH detection and showed great potential to detect AH in a large spatial extent.

1. Introduction

Abandoned houses (AH) are residential houses whose owners elected to give up their ownership, thus leading to long-term vacancy and obvious visual signs of vandalization (U.S. Government Accountability Office, 1978). In the U.S., housing abandonment and vacancy is a prominent national problem as the number of abandoned housing properties skyrocketed to 5.6 million in 2011, equal to around 5% of total housing units (U.S. Government Accountability Office, 2011; Mallach, 2018). As a result, AH have the most devastating impact on the urban environment, such as creating conditions for crimes, decreasing housing value in the neighborhood, and generating a substantial financial burden to local governments (Mallach, 2012; Raleigh and Galster, 2015; Molloy, 2016). To ameliorate urban decay and assist the revitalization of legacy cities, it is vital to derive the information of AH for better housing management in these neighborhoods (Silverman et al., 2013; Lynch and Mosbah, 2017).

Looking back on the history of housing abandonment studies, there was still a lack of a valid way to derive this information at fine resolution except labor-consuming field survey. Among the earliest efforts, housing abandonment was unveiled as a critical problem in the U.S. in 1978 with hopes of seeking a national solution to alleviate the situation (U.S.

* Corresponding author.
E-mail addresses: szou2@buffalo.edu (S. Zou), lewang@buffalo.edu (L. Wang).

https://doi.org/10.1016/j.isprsjprs.2021.03.020
Received 16 December 2020; Received in revised form 23 March 2021; Accepted 24 March 2021
Available online 3 April 2021

Government Accountability Office, 1978). However, since high foreclosure rates, population decline, and high unemployment contributed to the dramatically increased housing abandonment in certain cities, the situation aggravated (U.S. Government Accountability Office, 2011). In 2000, housing abandonment first raised serious concerns in academic communities as a significant problem that needed to be considered in urban development in the U.S. legacy cities (Accordino and Johnson, 2000). Since then, there were a significant number of urban studies modeling and predicting housing abandonment from various data sources, including socioeconomic data (Silverman et al., 2013; Morckel, 2014a; Bentley et al., 2016), nighttime light data (Yao and Li, 2011; Chen et al., 2015; Du et al., 2018; Wang et al., 2019), and geospatial data (Morckel, 2014b). In these studies, housing abandonment information was derived at relatively coarse spatial resolutions ranging from metropolitan area to census block. In contrast, fine-resolution urban studies in the U.S. legacy cities, e.g., small-area population estimation (Deng et al., 2010) and local-scale urban environment estimation (Frazier et al., 2013; Lynch and Mosbah, 2017), need individual-level housing abandonment as an important input factor. Such coarse resolution cannot meet the needs of urban studies at fine resolution. Besides, individual housing abandonment information is essential for government and real estate managers to make better decisions in urban planning and the housing market.

Despite the urgent need to detect individual-level AH, it has been barely explored with regard to a few existing publications. Among limited studies, three data sources were utilized to predict individual-level housing abandonment: field data (Scafidi et al., 1998; Hillier et al., 2003; Morckel, 2013; Yin and Silverman, 2015), utility data (Kumagai et al., 2016), and very-high-resolution (VHR) remote sensing data (Deng and Ma, 2015; Zou and Wang, 2020). As the most popular data source, field data, referred to housing information data that were collected through field survey, usually building information collected by governments, e.g., housing market value, floor area, and if it had been previously reported as vacant. As a representative study in recent years, Yin and Silverman (2015) utilized field data and geospatial data to predict properties' abandonment in Buffalo, NY. Multi-year models were employed in this study to better understand abandonment and demolition in the U.S. legacy city. Alternatively, Kumagai et al. (2016) employed utility data (hydrant data) in a basic Bayesian approach in housing abandonment modeling. The relationship between turned-off water hydrants and housing vacancy and abandonment was verified. Nevertheless, methods using either field data or utility data had a limitation when applied to a large geographic extension due to accessibility and data quality across different cities (Deng and Ma, 2015). To overcome this shortage, scholars also utilized remote sensing data in the detection by characterizing physical features of individual housing abandonment in two studies (Deng and Ma, 2015; Zou and Wang, 2020). The first remote sensing study integrated VHR remote sensing images and geospatial data to predict parcel-level housing abandonment and had over 80% accuracy (Deng and Ma, 2015). Later, Zou and Wang (2020) utilized VHR images as the major data source to explore the feasibility of remotely sensed physical features in AH detection. This exploration found that the prediction based on remotely sensed physical features had a limited producer's accuracy, and only vegetation conditions could infer housing abandonment (Zou and Wang, 2020). Thus, using remote sensing data to detect AH is not valid enough. As shown in

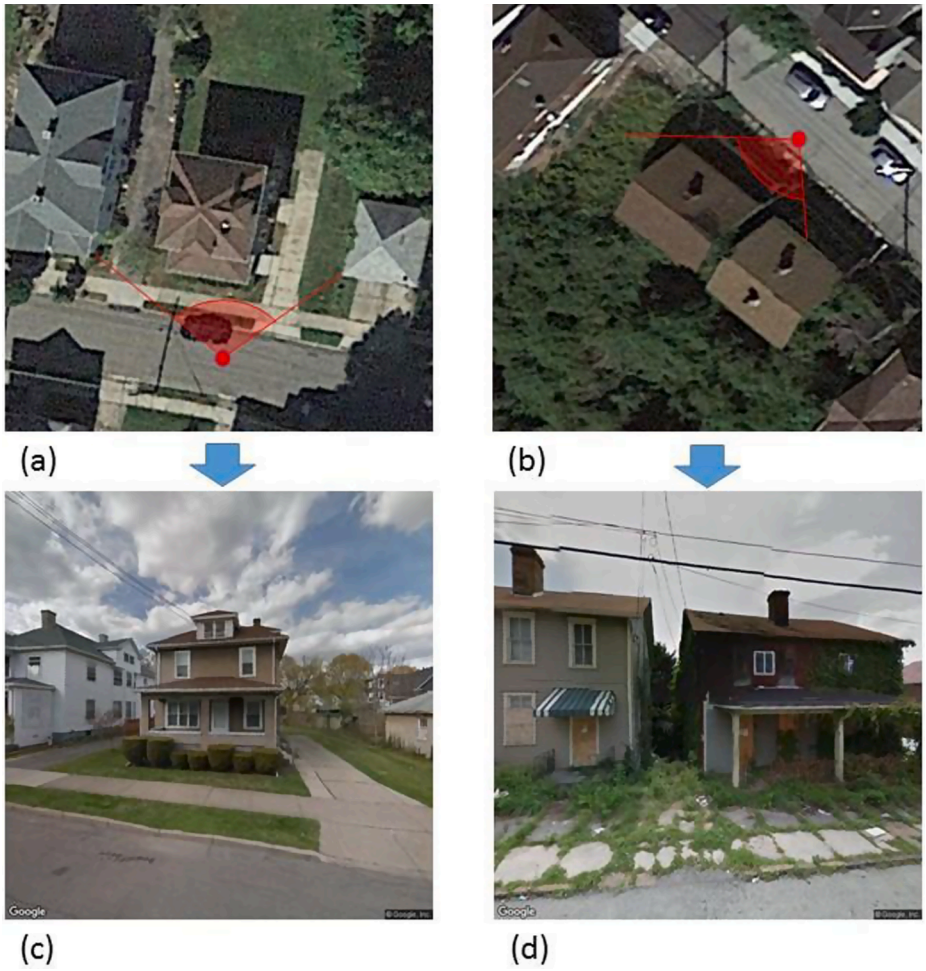

Fig. 1. Viewing OH and AH from remote sensing and street view images: (a) an OH in a VHR image; (b) a long-term AH in a VHR image; (c) the same OH in a street view image; (d) the same AH in a street view image. Red points and arcs in (a) and (b) show the viewing points and fields of view in street view images. An OH (a) and an AH (b) are not able to be differentiated from 0.3 m-resolution remote sensing images, while the same two houses have significantly distinctive appearances in street view images (c&d). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Fig. 1 a&b, traditional remote sensing data can only view from the top and provide information pertaining to the roof top and surrounding vegetation, which is not sufficient to differentiate AH with occupied houses (OH) (Zou and Wang, 2020). Considering the drawbacks in the above data sources, we need a new data source that is accessible and effective in AH detection.

Given the limitations in other data sources, Google Street View (GSV) imagery shows great potential for serving as a solution. Based on the comparison between remote sensing and GSV images of the same houses in Fig. 1, visual features of housing abandonment are better presented in the street view than in the top-view remote sensing. Thus, it is worthy of exploring the effectiveness of street view imagery. Also, among street view datasets, GSV is the largest and most representative one as it captures and provides millions of panoramic street view images daily at the global-scale (Anguelov et al., 2010). It has the broadest coverage, and covers more than 16 million kilometers of street view imagery across 83 countries until 2017 (Raman, 2019). GSV has become a cutting-edge data source in urban studies as the counterpart of remote sensing images, e.g., urban greenery assessment (Li et al., 2015), building use classification (Zhang et al., 2017; Li et al., 2017; Kang et al., 2018), and mobility pattern learning (Zhang et al., 2019). Therefore, the advantage of GSV includes both the broad coverage of remote sensing data and the street-level anthropomorphic perspective of field data, thus benefiting AH detection in effectiveness and generalizability.

A critical challenge arises when detecting AH with GSV images: how to develop a compatible method that can extract authentic the visual features of AH. This is due to the fact that GSV imagery contains excessive information, including both relevant (e.g., building façade) and irrelevant objects (e.g., sky, road). In general, existing deep learning methods were developed at two different scales, scene and patch, regarding the feature extraction of buildings. Scene-based methods utilized the whole street view scene as input in deep convolutional neural networks (CNN) to recognize the target object, which focused more on global features extracted from the image. For example, scene-based methods have been utilized to classify building functions (Kang et al., 2018), as well as estimating housing prices (Law et al., 2019). Alternatively, patch-based methods divided the whole scene into different homogeneous patches based on local features, followed by patch-level classification, which were ultimately integrated to make the scene-level decision. For example, patch-based methods have been proposed in building condition assessment (Koch et al., 2018) and building age estimation (Zeppelzauer et al., 2018). Regardless of the latest development of scene-based and patch-based methods, no efforts have been made to investigate if either method can be applied to AH detection. In a preliminary experiment, we observed that both global features (e.g., overall appearance of building facade) and local features (e.g., wooden blocked doors, broken eaves, and an overgrown lawn) could infer housing abandonment on the GSV image. However, how to combine the global and local features based on previously-developed scene or patch-based methods remains unsolved.

In summary, two distinctive gaps exist: The validity of street view image for abandoned house detection needs to be investigated; and, a compatible method needs to be developed to synthesize both the global and local features of abandoned houses. On account of these crucial gaps, we set aside two objectives in this study: (1) to explore the feasibility and accuracy of detecting individual-level abandoned residential housing units from GSV imagery; and (2) to develop a hierarchical deep learning approach to integrate global and local visual features in the detection.

2. Method

In this study, we developed a new hierarchical deep learning classification method to detect AH from GSV images (Fig. 2). After data

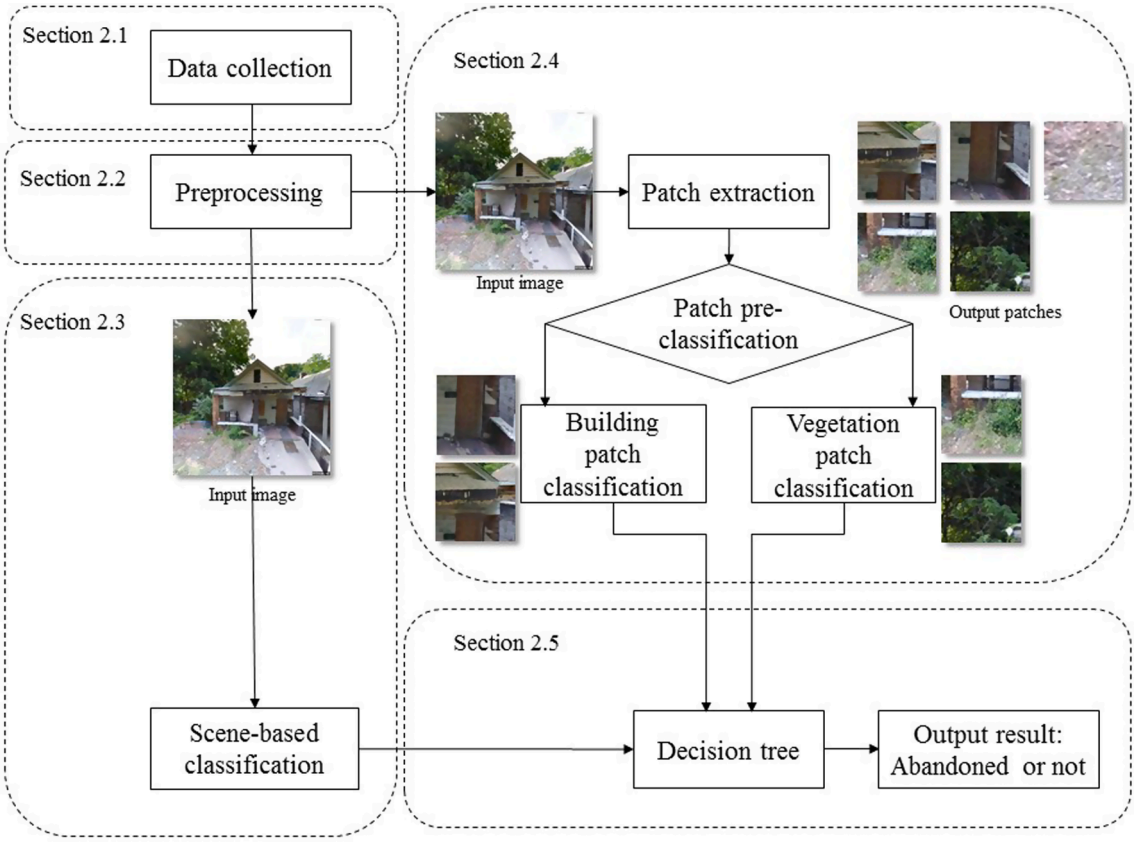

Fig. 2. Workflow of a hierarchical deep learning approach for AH detection using street view imagery.

collection and preprocessing, scene-based classification and patch-based classification were implemented simultaneously to extract global-level features and local-level features correspondingly. The scene-based classification (Section 2.3) utilized a pre-trained deep CNN model as the baseline in a transfer learning approach. In the patch-based classification (Section 2.4), patches were generated and then pre-classified into three categories: building, vegetation, and others. Using building and vegetation patches, two deep CNN models were fine-tuned to identify deteriorated building patches and overgrown vegetation patches independently. The proportions of identified patches in all patches were utilized as local-level features when integrating with the scene-based classification result in a decision tree model (Section 2.5). The final output result of the decision tree was individual-level housing abandonment status, i.e., whether the house within the GSV image was an AH or not. The detailed methodology was presented from Sections 2.1–2.5.

2.1. Image collection

There are two image datasets used in this work. The main dataset (denoted as Rust-Belt Dataset) covers five U.S. legacy cities across the Rust Belt region: Cleveland, OH, Chicago, IL, Pittsburgh, PA, Detroit, MI, and Binghamton, NY (Fig. 3). Cities in the Rust Belt region have suffered deindustrialization and depopulation since the middle of the 20th century, losing a significant number of manufacturing jobs and residents. The emigration of the population made this region the most representative concentration of AH (Griffin et al., 2015). For example, in Detroit, the population dropped by more than half from 1950 to 2015, and more than 1-in-3 properties, 139,699 of 384,672, suffered foreclosure because of mortgage defaults or unpaid taxes (Kurth and MacDonald, 2015). It is also notable that a demolition program was developed in these legacy cities to deal with the housing abandonment problem, e.g., the city of Detroit has demolished more than 20,000 houses as of the year 2020. Thus, the number of AHs in these cities is under a transition. Besides the dataset mentioned above (denoted as Rust-Belt Dataset in Table 1), we have an additional dataset (denoted as Detroit Test Dataset in Table 1) covering a distinct study site to test the effectiveness and demonstrate the implementation of the proposed method. This study site is located in a suburban area in Detroit, MI, and has no overlap with the Rust-Belt dataset. There are 340 parcels in this region (delineated in Fig. 7) located in a hyper-vacancy suburban area.

Table 1
Two image datasets in the experiment.

	Rust-Belt Dataset	Detroit Test Dataset
Coverage	Five cities across the Rust Belt region: Cleveland, OH, Chicago, IL, Pittsburgh, PA, Detroit, MI, and Binghamton, NY	A new site in Detroit, MI
#Addresses/ GSV images	2011	136
Training set (#images)	1000	0
Validation set (#images)	200	0
Test set (#images)	811	136
Notes	Spatially discrete distribution	Within a certain residential neighborhood

We collected raw images by retrieving and downloading street view images from the open-access GSV dataset. Two residential address datasets acted as input to collect potential AH and OH images, respectively. Abandoned single-family residential addresses were obtained from a commercial data company named Vacant House Data Feed (https://realestatewealthnetwork.com/vacant-house-data-feed/). In contrast, occupied single-family residential addresses were obtained from the OpenAddresses dataset (https://openaddresses.io/) after removing the addresses that overlap with the Vacant House Data Feed. Specifically, it should be noted that Vacant House Data Feed is a commercial dataset collecting vacant house addresses for housing managers, which not only includes houses that are abandoned but also houses that are temporally vacant and for sale. Therefore, the houses in the dataset including a large amount of well-maintained houses for sale are beyond the definition of abandoned houses making a manual labeling by visual analysis necessary to determine reliable ground-truth labels. Google Street View images for AH and OH were retrieved based on the above-selected addresses through GSV Static API (https://developers.google.com/maps/documentation/streetview/intro). For each house address in the address datasets, the API returned the photograph taken at the closest image-taking location facing the target address. The maximum

Fig. 3. The data set collection from five cities across the Rust Belt.

field of view was set as 120 degrees, and the degree of vertical angle of the camera was 0. The output size of the image in pixels was 640 × 640.

There were 18,964 GSV images collected as raw data based on addresses from Vacant House Data Feed and OpenAddresses datasets. A total of 7,580 (40%) of raw images were preserved after preprocessing. Among the 7,580 images, we sampled 2,011 images, including 1,007 AH images and 1,004 OH images for further process (Table 1). In the new test site in Detroit, MI, 136 of 340 parcels had residential properties and were available to detect AH.

2.2. Image preprocessing

Due to the uncontrolled quality of street view images, many raw GSV images cannot be used in the detection and need to be removed. These images have the following problems (see Fig. 4): (1) Occlusion. The target building is blocked by other objects, such as trees and vehicles, thus unable to appear in the image. (2) Misregistration. When capturing GSV images, the camera probably has a distant offset with the target house, resulting in the downloaded image not containing a complete residential property. A prerequisite of housing abandonment detection is that the GSV image must contain a clear and complete residential property in its central part. Either occlusion or misregistration can severely influence the detection results. Thus, we developed an image processing pipeline to remove these outliers.

When detecting and removing the outliers, we only reserved the central 320 × 320 pixels from downloaded GSV images (640 × 640 pixels) to focus on the central part of the image and the target building. Then, referring to a previous study (Kang et al., 2018), we employed a ready-made scene recognition deep CNN model, the VGG16 model (Simonyan and Zisserman, 2014) trained on the Places365 dataset (Zhou et al., 2017), to filter the cropped street view images. Based on the classification scheme in Places365, we selected nine target classes: [apartment, balcony-exterior, house, beach house, barn, building facade, cabin-outdoor, cottage, doorway-outdoor]. Only the cropped images belonging to the target classes were preserved for the following approach. A total of 7,580 (40%) raw images were reserved and 11,384 (60%) downloaded raw GSV images were filtered out. Finally, we selected 1,007 abandoned houses and 1,004 occupied houses as samples by visual interpretation based on their appearances on the GSV images. This step was to collect GSV image samples of AH and OH more confidently, thus ensuring the purity of the training set.

2.3. Scene-based classification

In this section, we implemented a transfer learning approach in the scene-based classification to extract global features from street view images. GSV images that were reserved in Section 2.2 were utilized as inputs to train the CNN model directly. We labeled the housing abandonment status of each GSV image through visual interpretation. Images were randomly divided into about 50% for training (1000 images), about 10% for validation (200 images), and about 40% for test (811 images). We implemented a transfer learning strategy (Weiss et al., 2016) to fine-tune a pre-trained CNN model instead of training from scratch. A deep CNN model, VGG16, that had been pre-trained on the Places365 dataset was fine-tuned through the learning approach. Images were resized to fit the input of the CNN model before learning. The output of this binary classification was the predicted category and also the probability of the scene being in two categories. We utilized the probability of being in the AH class to represent the scene-based classification results in the final decision-tree model.

The learning approach was implemented in Caffe (Jia et al., 2014), operated on Linux Ubuntu 16.04, and carried out by one NVIDIA GTX 1080 Ti GPU, which is accelerated by cuDNN5. Training was performed for 50 epochs with a batch size of 10, a learning rate of $1 * 10^{-3}$, a momentum of 0.9, and a decay of 0.1 in learning rate every 10 epochs. Cross-entropy loss was utilized for training with the weight decay parameter $5 * 10^{-4}$.

2.4. Patch-based classification

In a GSV image, objects may appear in different scales, perspectives, lighting conditions, and locations. A patch-based method can help eliminate noise in these situations and focus on local features at multi-scales (Koch et al., 2018). Note that patch here refers to a square-shaped connected region of the image that contains one feature point at the center and its neighborhood pixels. In visual interpretation, there are two observable characteristics on street view images that enable one to distinguish between AH and other houses: a deteriorated building facade and overgrown vegetation. On account of long-term abandonment, these houses reveal significantly worse building conditions that include deteriorated facade and outer walls, damaged facade by vegetation, or wooden blocked windows and doors. Building condition measures the effectiveness of current maintenance programs (Abbott et al., 2007). Alternatively, vegetation condition refers to overgrown lawns, shrubs, or trees in front yards due to long-term abandonment and

Fig. 4. Occlusion and misregistration in GSV images.

out of maintenance. The intensity of vegetation overgrowth in an AH is supposed to be more severe than a lack of mowing in a short period of time, sometimes even damaging the buildings. The knowledge about these two characteristics, which was gained from visual interpretation, guided the design of the patch-based deep learning approach. In terms of two aspects of visual features, the proposed patch-based classification focuses on building and vegetation conditions separately. Patches were first automatically extracted from GSV images. Then, a content-based retrieval approach was developed to eliminate irrelevant patches and focus on building and vegetation conditions. Lastly, patches were classified in deep CNN models to identify deteriorated building patches and overgrown vegetation patches.

2.4.1. Patch extraction and pre-classification

Patch extraction includes the following steps. We extracted the pyramid histogram of visual words (PHOW) (Bosch et al., 2007) features at three static scales from each image. The PHOW features are a variant of dense scale-invariant feature transform (SIFT) descriptors, which is a commonly used feature detection algorithm to fast detect and describe local features in images. This step generated tens of thousands of feature keypoints and their feature descriptors at three scales based on three pre-determined bin sizes (Bentley et al., 2016; Deng et al., 2010; Griffin et al., 2015). A descriptor vector was associated with each feature keypoint. To reduce the redundant features and speed up the follow-up approach, we filtered out features with low contrast, which referred to the norm of the descriptor vector. Then, k-means clustering ($k = 39$) was employed for all remaining feature descriptor vectors in a given image. For each cluster, 1% features that are nearest to the cluster centroid were extracted. In the clustering, the Euclidean distance between feature descriptor vectors and the cluster centroid descriptor vector was employed as the measure. Next, one of the extracted features that had the highest contrast in a cluster was selected as the representative for each cluster. Thus, the number of PHOW features was reduced to 39, each of which represents one cluster. Based on the selected feature keypoints' location and scale, square-shaped patches were cropped from the image in multiple sizes. Specifically, the location of patch center was the feature keypoint and the size of patch was determined through multiplying the bin size of the feature with a constant, 16. Therefore, patches were extracted based on high-contrast SIFT features, which omitted regions with low edge contrast regions. Last, to better present vegetation condition, which has lower edge contrast, a fixed-size patch usually containing the front-yard lawn, which locates at the central bottom image, was cropped from each image. Adding 39 PHOW feature patches, we extracted a total of 40 patches from each image. This patch extraction approach enabled us to automatically extract local features at multiple scales eliminating the disturbance from low-contract regions, e.g., sky. At the same time, clustering and the following steps reduced the number of features and effectively selected representatives from similar features.

When building facades and vegetation were represented densely and completely in the patches, irrelevant and redundant patches, e.g., cars, driveways, pedestrians, and traffic signs, still exist. To remove these irrelevant patches and identify building and vegetation patches, we pre-classified patches into three categories: building, vegetation, and other objects using a CNN model, VGG16 trained on Places365, with pre-determined target classes. We defined 42 Places365 categories as target classes for building: [alley, garage-outdoor, elevator lobby, hangar-outdoor, schoolhouse, house, arch, elevator door, fire escape, courthouse, mansion, manufactured home, mausoleum, oast house, atrium, attic, porch, ruin, sauna, server room, shed, skyscraper, staircase, tower, balcony exterior, balcony interior, barn, barn door, basement, beach house, bow window-indoor, building facade, clean room, chalet, apartment building-outdoor, cabin-outdoor, campus, castle, doorway-outdoor, embassy, office building, parking garage-outdoor], and thirteen categories as target classes for vegetation: [yard, tree farm, field-wild, field-cultivated, forest-broadleaf, greenhouse-indoor, lawn, orchard, forest path, cornfield, marsh, rice paddy, bamboo forest]. We consulted both the semantic meaning of the target class and the illustrated photographs in the Places365 dataset (http://places2.csail.mit.edu/explore.html) when selecting the target classes. The target classes in the Places365 should only contain building (or vegetation) elements in the dataset. Also considering the various styles of residential building facades, we attempted to include as many related classes as possible. Patches belonging to the target classes were preserved for further steps.

2.4.2. Patch classification

A deep learning classification approach was employed to identify deteriorated building patches and overgrown vegetation patches in this section. Building patches and vegetation patches were classified independently. Building patches were classified into two categories: good condition and deteriorated condition, and vegetation patches were classified similarly into two categories: good condition and overgrowing condition. All patches were manually labeled by visual interpretation and resized before classification to fit the input of the CNN model. Patch labeling was independent of the scene labeling. Even extracted from a single scene, patches are in very different conditions. For example, patches extracted from an AH image may not present any sign of deterioration, thus cannot represent the abandonment status of the house. Considering the fact that patch labels may be inconsistent with the scene label, automatic patch labeling by propagating the label assigned to the image/scene to the patches is infeasible. Patch labeling still needs manual input. As mentioned in the scene-based method, images were divided thusly, 50% for training, 10% for validation, and 40% for test. Based on which image they belong to; patches were partitioned into the three sets, respectively. Thus, we chose the same pre-trained CNN model as the baseline to fine-tune in both the building and vegetation patch learning approaches. The classification result of each model was the binary condition of a patch. Thus, the deteriorated building patches and the overgrown vegetation patches can be identified.

Three representative deep CNN architectures pre-trained on the Places365 dataset: AlexNet (Krizhevsky et al., 2012), VGG16 (Simonyan and Zisserman, 2014), and ResNet152 (He et al., 2016) were tested in a preliminary attempt. These three CNN models showed micro-close performance in the learning approach and test results. Thus, in the following approach, only the learning approach and results using VGG16, which had the moderate model complexity and runtime, was implemented and presented in this study. In the training and test approach in patch-based classification, we used the same setup with the scene-based classification.

2.5. Integrating scene-based and patch-based classification results in abandoned house detection

We developed a decision-tree model to integrate the scene-based classification results and the patch classification results in individual-level housing abandonment prediction. First, within each image, patch-wise classification results were used to calculate the proportion of deteriorated building patches in building patches, $BPR_{building}$, and the proportion of overgrown vegetation patches in vegetation patches, $BPR_{vegetation}$. In Eq. (1), BPR is the ratio between bad-condition building (or bad-condition vegetation) patches to all building (or vegetation) patches for each image, which is calculated by dividing the number of deteriorated building (or overgrown vegetation) patches, BP, by the number of all building (or vegetation) patches in a certain image, AP.

$$BPR_p = \frac{BP_p}{AP_p}, \quad \text{p is vegetation } or \text{ building} \tag{1}$$

Then, we generated a decision tree based on three variables, $BPR_{building}$, $BPR_{vegetation}$, and the probability of AH, which was the scene-based classification result. $BPR_{vegetation}$ and $BPR_{building}$ represented local-level features of AH, while the probability of AH from the scene-based classification represented the global-level features. The hierarchical

features were integrated in a decision tree to predict the housing abandonment at the individual level. The decision tree was fitted by the training dataset and employed to predict the housing abandonment in the test dataset.

3. Results

3.1. Scene-based classification results

The learning approach of scene-based classification is presented in Fig. 5. In the first several epochs, validation accuracy increased dramatically. Later with the learning proceeding, the curve fluctuated and became constant at around 83%. Changes in validation and training loss verified this trend, decreasing in the beginning, then fluctuating and stabilizing. In the test result (Table 2), the overall accuracy achieved 80.4%, which was slightly lower than validation accuracy. All producer's accuracy and user's accuracy were 81.1% and 80.1%, making the F-score 0.80. The kappa coefficient was 0.60, indicating a moderate agreement between the prediction and ground truth.

3.2. Patch-based classification results

3.2.1. Patch extraction and pre-classification

In section 2.4, we extracted 80,440 patches from 2,011 images. Building facades and vegetation were represented densely and completely in the patches (Fig. 6). Among these patches, 45,559 patches were obtained after pre-classification, including 34,208 building patches and 11,351 vegetation patches. The pre-classification accuracy was over 98% from visual estimation on a random sample set which includes 1,000 patches.

3.2.2. Patch classification

We had 17,211 building patches and 5,731 vegetation patches for 1,000 images in the training set, 3,388 building patches and 1,113 patches for 200 images in the validation set, and 13,609 building patches and 4,507 vegetation patches for 811 images in the test set. Training approaches for both building and vegetation patches presented similar trends in the change of accuracy and loss (Fig. 7). With the learning proceeding, validation accuracy increased significantly at the very beginning, from below 30% to over 80%, then tended to be more constant in further epochs. This indicates the models converged in several epochs after the beginning of learning. The training loss (red lines in Fig. 7b&d) kept reducing in learning, while validation loss (blue lines in Fig. 7b&d) increased in the first 20 epochs and tended to be constant after 20 epochs. These trends indicate that the model presented overfitting after convergence. The early convergence is probably due to the very relevant pre-trained model used in this transfer learning approach. An optimal model for the latter approach was selected.

Table 2
Confusion matrix of AH detection using scene-based classification.

		Ground Truth			
		AH	OH	Total	User's accuracy
Predicted	AH	330	82	412	80.1%
	OH	77	322	399	80.7
	Total	407	404	811	
	Producer's accuracy	81.1%	79.7%		Overall: 80.4%

*Kappa = 0.60, indicating a moderate agreement. F-score = 0.80.

Among 13,609 building patches in the test, 11,389 were correctly classified, making the overall accuracy as high as 83.7% (Table 3). However, the producer's accuracy of bad-condition building patch detection was low (44.1%), which indicated that a large proportion of the deteriorated building patches were misclassified, thus placing them in the wrong class. The same situation occured in vegetation patches when the overall accuracy was 86.3% (Table 4) and the producer's accuracy of bad-condition vegetation patch was as low as 54.4%. Thus, the omission errors, 55.9% and 45.6%, for both bad-condition building and vegetation were considerable.

3.3. Individual-level housing abandonment result

As shown in Table 5, integrating scene-based and patch-based results, the developed decision-tree model predicted housing abandonment with 85.0% overall accuracy and 0.70 kappa coefficient. The producer's accuracy of AH was 77.4%, and the user's accuracy was 92.6%, making the F-score 0.84. Compared with the baseline scene-based classification model, the hierarchical model had higher overall accuracy as well as the F-score and the kappa coefficient (Table 6). The

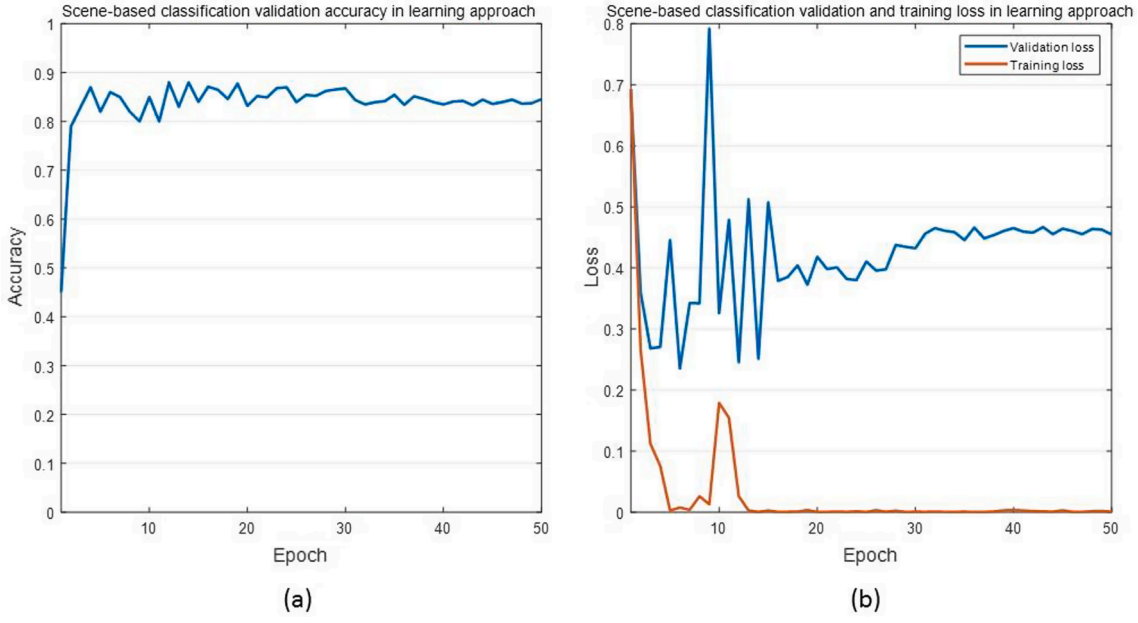

Fig. 5. The learning curve of scene-based classification: (a) validation accuracy in learning approach; (b) validation loss (blue) and training loss (red) in the learning approach. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

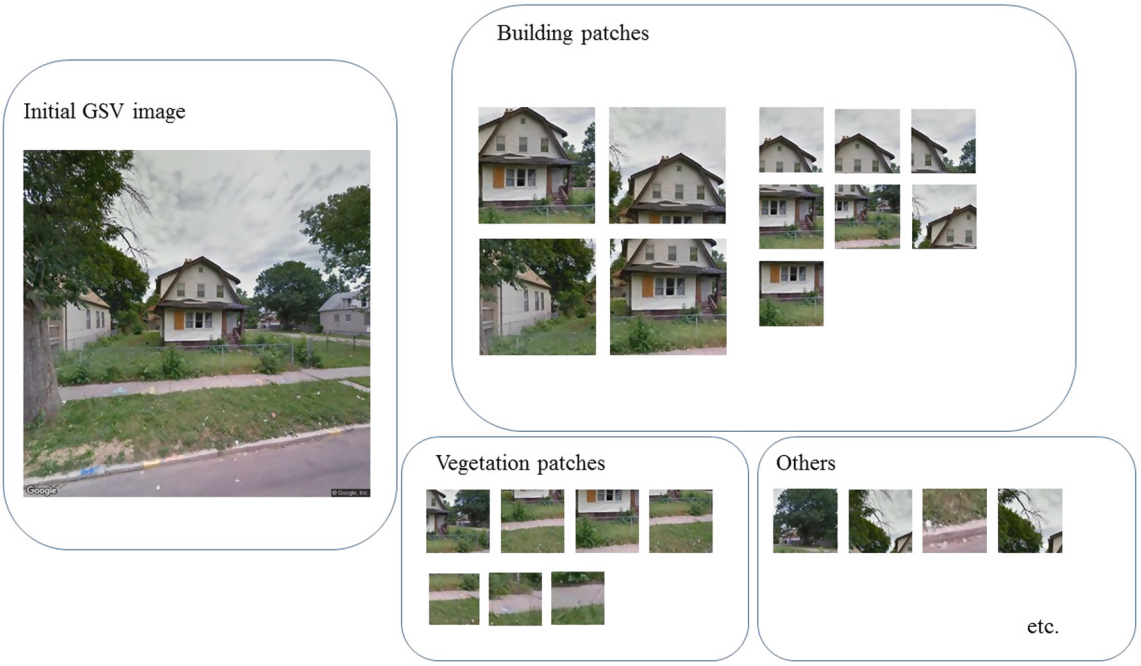

Fig. 6. An example of a GSV image and its building patches and vegetation patches.

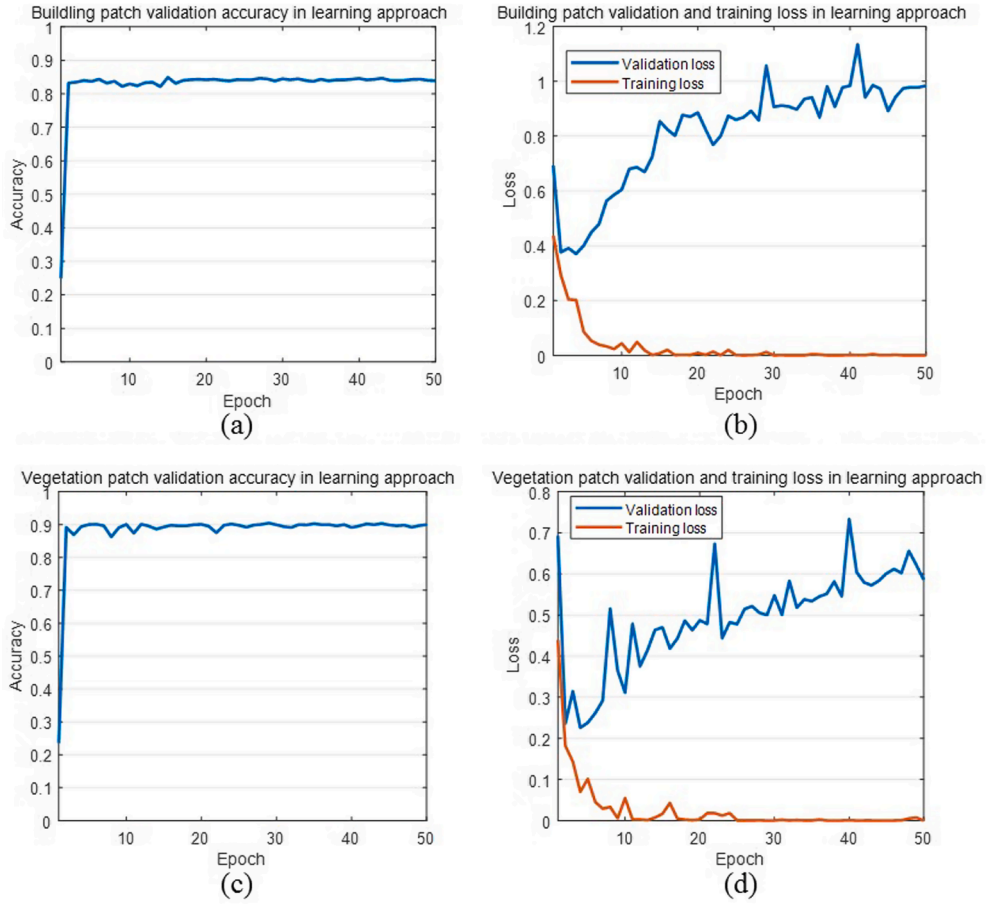

Fig. 7. The learning accuracy and loss curves: (a) validation accuracy in building patch learning; (b) validation loss (blue) and training loss (red) in building patch learning; (c) validation accuracy in vegetation patch learning; (d) validation loss (blue) and training loss (red) in vegetation patch learning. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 3
Confusion matrix of building patch classification using VGG-16.

Building Patch		Ground Truth			
		Bad condition	Good Condition	Total	User's accuracy
Predicted	Bad condition	1345	515	1860	72.3%
	Good condition	1705	10,044	11,749	85.5%
	Total	3050	10,559	13,609	
	Producer's accuracy	44.1%	95.1%		Overall: 83.7%

*Kappa = 0.46, indicating a moderate agreement. F-score = 0.55.

Table 4
Confusion matrix of vegetation patch classification using VGG-16.

Vegetation Patch		Ground Truth			
		Bad condition	Good Condition	Total	User's accuracy
Predicted	Bad condition	631	89	720	87.6%
	Good condition	529	3258	3787	86.0%
	Total	1060	3347	4507	
	Producer's accuracy	54.4%	97.3%		Overall: 86.3%

*Kappa = 0.59, indicating a moderate agreement. F-score = 0.67.

Table 5
Confusion matrix of AH detection integrating scene-based classification and patch-based classification.

		Ground Truth			
		AH	OH	Total	User's accuracy
Predicted	AH	315	30	345	91.3%
	OH	92	374	466	80.3%
	Total	407	404	811	
	Producer's accuracy	77.4%	92.6%		85.0%

*Kappa = 0.70, indicating a substantial agreement. F-score = 0.84.

Table 6
Accuracy comparison between Hierarchical AH detection method and scene-based AH detection.

Methods	Overall accuracy	Producer's accuracy	User's accuracy	F-score	Kappa
Scene-based	80.4%	81.1%	80.1%	0.80	0.60
Hierarchical	85.0%	77.4%	91.3%	0.84	0.70

user's accuracy had significant improvement from 80.1% to 91.3%, while the producer's accuracy slid slightly from 81.1% to 77.4%.

3.4. Mapping AH detection results in Detroit, MI

Since the hierarchical model had better performance than the scene-based method, we chose this model to apply in generating a map of the AH detection result. A map of a new study site in Detroit, MI, was generated to examine the feasibility and accuracy of the proposed model (Fig. 8). In this study site, 136 of 340 parcels have residential properties and available street view images for detection. In the detection result, a total of 108 test houses were correctly classified, making the overall detection accuracy achieve 78.7% (Table 7). The kappa coefficient was 0.56, indicating a moderate agreement with ground truth. Forty AH were correctly detected, while twenty OH were wrongly detected as AH, and nine AH were missing in the detection. Accuracies of the test in this new study site were lower than in the original dataset but still acceptable. Two hundred and four parcels that were not included in the detection were labeled as *Others* in Fig. 8.

4. Discussion

To our knowledge, this is the first attempt to demonstrate the feasibility of street view imagery in AH detection. To effectively leverage global features and local features in the detection, a new hierarchical deep learning approach was developed and tested. Compared with the benchmark scene-based method, the hierarchical method can effectively improve the user's accuracy. Both data source and method in this paper are revolutionary in AH detection. Advantages, limitations, and further studies are discussed in the following.

4.1. Contributions and advantages

4.1.1. Google street view

There are two advantages of using GSV imagery in AH detection. First, street view imagery provides unique street-level information that is effective in identifying housing abandonment. In this study, the effectiveness of GSV was demonstrated in the experiments. Abandoned house detection using street view imagery as a single data source achieved 85% in overall accuracy and 0.84 in F-score. In addition, the test in a new study site has 78.7% overall accuracy and a 0.74 F-score. As shown in Table 8, these accuracies are comparable with previous detection studies using other data sources (Yin and Silverman, 2015; Deng and Ma, 2015; Kumagai et al., 2016) and significantly better than the previous study using remote sensing data only (Zou and Wang, 2020). Field survey data is considered the most reliable data source to predict housing abandonment when the F-score is as high as 0.98 (Yin and Silverman, 2015). Another fact is that previous studies usually involved multiple data sources in the detection. The previous study integrating utility data and building information from field surveys can also achieve a 0.88 F-score (Kumagai et al., 2016). As low-cost and open-access data sources, geospatial data, and remote sensing data were also employed and had a 0.80 F-score (Deng and Ma, 2015). When using remote sensing data only, the F-score is 0.49 (Zou and Wang, 2020). Regarding accuracy, GSV is more effective than VHR remote sensing imagery in AH detection and is comparable with other data sources in the detection. It should be noted that utility data is an accurate and timely data source according to the previous study (Kumagai et al., 2016). When utility data is accessible, e.g., for local policymakers and governments, detecting housing abandonment using utility data is ideal. However, for researchers and commercial housing managers, it is not easy to obtain data from utility companies due to privacy protection and interest conflict. In addition, when making a decision for a large region, the use of utility data may need data from multiple utility companies, thus the quality of data may be inconsistent. As an open-access dataset with a large coverage, GSV images could be an auxiliary or even an alternative data source in AH detection in large-scale decision making and for research and commercial use.

Another advantage is that GSV has good accessibility and generalizability for a large-regional study. As mentioned in the Introduction, GSV imagery is an open-access dataset that has broad coverage throughout the world, thus having good accessibility. Besides, the effectiveness of GSV is more consistent in a large region than other data sources. Previous modeling studies based on socioeconomic data have different statistically significant variables in different study sites (Wiechmann and Pallagst, 2012; Morckel, 2014b; Yin and Silverman, 2015), and the extent of previous studies usually limited in a city (Deng and Ma, 2015; Kumagai et al., 2016). The spatial heterogeneity of significant variables makes the models based on socioeconomic and built-environment data may perform differently in different study sites. Subsequently, the effectiveness may be inconsistent. Alternatively, the visual pattern of AH is a more consistent feature that makes AH

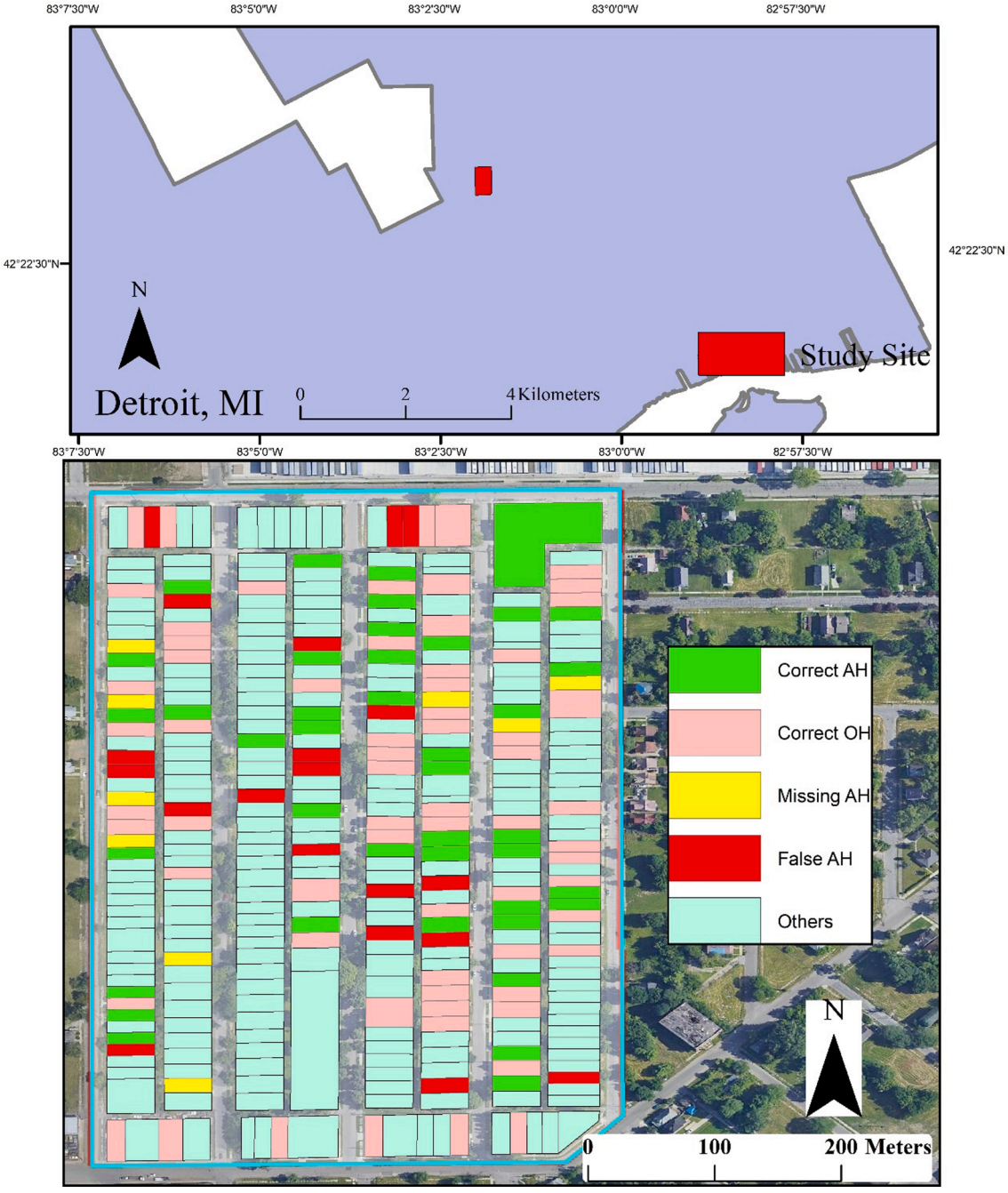

Fig. 8. AH detection in Detroit, MI. The F-score is 0.74.

Table 7
Confusion matrix of AH detection in Detroit, MI.

		Ground Truth			
		AH	OH	Total	User's accuracy
Predicted	AH	41	20	61	67.2%
	OH	9	66	75	88.0%
	Total	50	86	136	
	Producer's accuracy	82.0%	76.7%		Overall: 78.7%

*Kappa = 0.56, indicating a moderate agreement. F-score = 0.74.

Table 8
Accuracies of abandoned house detection studies using different data sources.

Data source	Field survey data and geospatial data	Utility data field survey data	VHR remote sensing image and geospatial data	VHR remote sensing image	GSV image
F-score	0.98	0.88	0.80	0.49	0.85
Reference	Yin and Silverman, 2015	Kumagai et al., 2016	Deng and Ma, 2015	Zou and Wang, 2020	This paper

distinguishable. We presented prediction accuracy across five representative cities in the Rust Belt region in Table 9. Based on the results, all overall accuracies for cities are over 78%, while the lowest overall

Table 9
Detection accuracies in five cities across the Rust Belt region.

City	Pittsburgh, PA	Cleveland, OH	Chicago, IL	Binghamton, NY	Detroit, MI	New site in Detroit, MI
Overall accuracy	83.6%	86.7%	79.4%	92.6%	91.7%	78.7%
#Test samples	397	128	126	95	70	136

accuracy is still acceptable (78.7%). Four study sites have over 83% overall accuracy (83.6–92.6%), while Binghamton, NY has the highest 92.6%; and, Chicago, IL and the new test site in Detroit, MI have the lowest accuracies. This spatial diversity in accuracy is mainly due to the difference in the proportion of AH in different regions. The producer's accuracy (Table 5) of OH is significantly higher than AH. The prediction accuracy in hyper-vacancy regions, twill be lower than other regions. Nevertheless, considering these cities are from five different states, the proposed model based on GSV has relatively consistent performance across the Rust Belt. Visual features in GSV are considered as a more consistent indicator of housing abandonment than socioeconomic and built-environment data. As far as we are concerned, thanks to its accessibility and consistency, GSV imagery is the most promising data source to realize a nationwide AH detection.

4.1.2. The hierarchical deep learning approach

Compared with the traditional scene-based method, the proposed hierarchical approach presented breakthroughs. The hierarchical deep learning approach better simulated the process of visual interpretation than the scene-based method by identifying both global and local visual features. In the experimental results, the scene-based classification achieved 80.4% overall accuracy. Based on the result, we inferred that global-level visual features in GSV images could indicate housing abandonment. For example, long-term abandoned houses usually have overgrown plants climbing on the building, which will result in overall characteristics within the image, such as a high proportion of vegetation and abundant texture information. Furthermore, the hierarchical deep learning approach had better performance than the scene-based method owing to the significant improvement in the user's accuracy, from 80.1% to 91.3%. Specifically, when using a hierarchical method, more than 60% of wrongly detected abandoned houses were reduced (from 82 to 30), while only eleven AH were missed in the detection and the producer's accuracy decreased 3.7%. This indicates that local features can make the detection more accurate. A possible reason is that when some of OH and AH have similar overall appearance, the proportion of bad-condition building and vegetation patches, which reflect local features, can quantitatively assess the deterioration and facilitate detection. Moreover, the proposed patch-based classification can focus on local features in specific categories by eliminating irrelevant objects. The patch extraction and pre-classification excluded low-contrast information and irrelevant patches effectively. In feature extraction, we kept high-contrast feature keypoints and extracted patches based on these feature points. Low-contrast irrelevant elements, e.g., sky, were filtered out. In patch pre-classification, we classified patches into three categories: building, vegetation, and others. The category others (cars, roads, sidewalks, pedestrians, traffic signs, etc.) were excluded. Then, building and vegetation patches were labeled and classified separately. When labeling the patches, we labeled the relevant patches that indicate housing abandonment as 1, and others were 0. This strict labeling criterion reduces the impact from irrelevant elements, like trees, to the prediction, since we utilized the proportion of bad-condition patches (labeled as 1) as an AH indicator. Only using building patches or vegetation patches to detect AH achieved 80.9% and 68.8% overall accuracy separately. The integration result of building patches and vegetation patches showed 81.3% in overall accuracy, which is slightly better than the scene-based classification but significantly worse than the hierarchical results. This preliminary result indicated that both local features benefited the detection, and building façade condition was relatively more important than vegetation condition. By integrating authentic global and local features, the hierarchical approach empowers deep CNN substantially and has great potential in further street-view urban studies.

4.2. Uncertainty and limitations

4.2.1. Training and test data uncertainty

Resulting from manual sample selection, errors in training data are ineluctable. In sample selection, only AH and OH with very high confidence in visual interpretation were selected as samples to ensure the purity of the samples. Therefore, this data collection method may involve bias in the dataset, e.g., only AH that have an extremely deteriorated facade being selected and the vegetation condition being overestimated. These biases may result in both limited AH samples and false AH samples in the training dataset, consequently leading to false and missed detection in the results. When we applied the proposed model in a new dataset that covered all samples in a region, the overall accuracy decreased from 85.0% to 78.7% (Table 7). It is likely that the developed model based on the manually selected training data can only identify obvious deterioration or severely overgrown vegetation; and, thus, cannot handle the scenario when AH and OH have similar appearances. Subsequently, this study performed better in detecting long-term AH with obviously bad conditions than other AH. Nevertheless, visual interpretation is the most efficient way to collect reasonable AH samples on a large scale considering that AH datasets are various or even absent in different cities, and the definition of AH is vague between datasets and studies (Morckel, 2014a; Mallach, 2018). For further study, a nationwide AH dataset with a universal definition is needed. Surging volunteered geographical information provides the opportunity to create a nationwide AH dataset through public participation and contribution in providing data voluntarily (Elwood et al., 2012).

4.2.2. Limitations in Google street view imagery

Limitations in GSV data can contribute to errors in the detection and impede AH mapping. First, adjacent buildings disturb the detection result as the image of the target building may contain adjacent buildings and front yards. For example, in patch-based classification, irrelevant building and vegetation patches can still be involved in the detection. More specifically, when the target house is well-kept and its adjacent house is abandoned, the target house will be wrongly identified as an AH because of bad-condition patches on the adjacent house. This problem resulted in falsely detected AH, and thus reduced the user's accuracy, especially when using a test dataset not selected manually. This partially explains why the user's accuracy is as low as 67.2% in the new test site in Detroit. A potential solution is identifying each house from GSV images first, then geolocating each structure to coincide with their target addresses. Laumer et al. (2020) proposed a global optimization approach to geolocate trees detected in GSV images. Although this method has not been tested to geolocate buildings, it provides a potential solution to accurately locate the target building in the GSV image before detection.

Time inconsistency and occlusion in street view data restrain the mapping of individual-level AH. As mentioned in Li et al.'s paper (2015), even within a small area, only a few images were captured simultaneously. The time difference can be ten years, which means a detected AH that existed in 2009 may disappear in 2019 due to demolition or reoccupation. Therefore, generating a unified-time map of AH using GSV imagery is unachievable. The time inconsistency in street view imagery needs to be considered when further implementing the proposed model in a regional study. In addition, in order to handle the uncontrolled

quality of GSV images, we filtered out images where houses were occluded by trees or vehicles and labeled them as blanks in the detection. As shown in Fig. 8, the class *Others* includes a significant number of residential properties that do not show up on GSV images due to occlusion and misregistration. With these blanks in the detection, it is difficult to detect and map every AH in a given region. In this study, as training and test samples were collected discretely and preprocessed, this problem was not a concern in the classification. Kang et al. (2018) proposed a feasible way to alleviate this problem by utilizing the three closest GSV images of the target building instead of only the closest one, which increased the probability of covering every house within a region. This strategy can be applied in further studies to generate a better map. Furthermore, Google provides multi-temporal GSV images for some regions on Google Maps. These images cannot be bulk downloaded through Google Street View API yet. Multi-temporal GSV images can provide nearly up-to-date information and improve the classification by fusing multi-temporal information at the image level, the feature level, or the decision level and by helping to eliminate time-sensitive irrelevant objects. Also, multi-temporal GSV may provide observations from different viewing aspects, hence, overcoming occlusion limitation. This potential extends the application value of the proposed method. In addition, multi-temporal observations on abandoned houses will help to understand the formation, deterioration, and demolition of abandoned houses. Once this dataset becomes available for bulk download, it has great potential in improving AH detection and other urban applications, such as quality of street space estimation (Li and Long, 2019).

4.2.3. Limitations in the hierarchical approach

Limitations exist in the hierarchical approach, especially in the patch-based method. First, patch labeling is time consuming and difficult. As the number of patches greatly exceeded the number of images, patch labeling was labor-consuming. Also, some patches are too small for class recognition. Thus, we labeled these small-sized and confusing patches as good-condition patches, which may cause underestimation of AH. Second, the performance of the proposed method depends on the parameters in the patch-based method. For example, one of the important parameters is the number of extracted patches. We traded off the count of resulting patches and the expression of features to determine a reasonable value for the number of extracted patches. If we extracted too many patches, they would overlap heavily, and more labor is needed in manual labeling, while if we extracted too few patches, essential features could be missed in the resulting patches. At the same time, we need to ensure authentic features can be extracted from all GSV images with a sufficient number of patches. Thus, the number of patches can be modified based on different patch sizes and specific features in further studies. Last, patches were in various styles even when they were extracted from one house image. This diversity brought great difficulties in the classification. In addition, a patch is not a completely homogeneous region. Even many irrelevant objects have been eliminated in feature extraction and patch pre-classification, there were still disturbed objects within a patch. This problem was inevitable if the classification was implemented based on the current connected square-shaped regions in such a complex urban environment, because the exact boundary of objects was not delineated clearly. An ideal method to accurately extract featured regions is the semantic segmentation method (Lin et al., 2020), which is expected to be more time consuming in labeling than the proposed method.

4.3. Further study

Beyond the above mentioned further directions, further studies may also focus on buildings for other uses (e.g., commercial and industrial buildings), integrating more data sources in the detection or applying the detection results in fine-resolution urban studies for further application. Integrating multi-source data in the detection, including GSV data, remote sensing data, demographic data, or geospatial data, is a potential direction to explore the improvement of accuracy. Also, the individual-level AH dataset derived from our method can play as input variables in various fine-resolution urban studies, e.g., small-area population estimation, built environment estimation.

5. Conclusions

Unlike traditional remote sensing data, street view images provide brand-new information of building facade structure and front-yard vegetation status, both of which are essential in AH detection. In this paper, we demonstrated the effectiveness of GSV imagery in individual-level AH detection and proposed the first deep learning approach for the detection. The new data source and the new hierarchical deep learning approach empower the developed model to achieve substantial accuracy. Specifically, GSV imagery, as a new data source, is outstanding in effectiveness and accessibility in AH detection and performed consistently across the Rust Belt region. Also, the proposed hierarchical deep CNN approach that can integrate global and local features has great potential in ground-based image processing in urban studies. Regarding the advantages and limitations in data and method, generalizing the proposed method to a large region is cautiously optimistic. The detection results will aid the “Smart City” initiative by helping better understand the urban environment and bringing the new opportunity to high-resolution urban studies, housing management, and urban planning.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Abbott, G.R., McDuling, J.J., Parsons, S.A. and Schoeman, J.C., 2007. Building condition assessment: a performance evaluation tool towards sustainable asset management.
Accordino, J., Johnson, G.T., 2000. Addressing the vacant and abandoned property problem. J. Urban Affairs 22 (3), 301–315.
Anguelov, D., Dulong, C., Filip, D., Frueh, C., Lafon, S., Lyon, R., Ogale, A., Vincent, L., Weaver, J., 2010. Google street view: Capturing the world at street level. Computer 43 (6), 32–38.
Bentley, G.C., McCutcheon, P., Cromley, R.G., Hanink, D.M., 2016. Race, class, unemployment, and housing vacancies in Detroit: an empirical analysis. Urban Geography 37 (5), 785–800.
Bosch, A., Zisserman, A., Munoz, X., 2007, October. Image classification using random forests and ferns. In: 2007 IEEE 11th International Conference on Computer Vision. IEEE, pp. 1–8.
Chen, Z., Yu, B., Hu, Y., Huang, C., Shi, K., Wu, J., 2015. June. Estimating house vacancy rate in metropolitan areas using NPP-VIIRS nighttime light composite data. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 8 (5), 2188–2197.
Deng, C., Ma, J., 2015. Viewing urban decay from the sky: A multi-scale analysis of residential vacancy in a shrinking US city. Landscape Urban Plann. 141, 88–99.
Deng, C., Wu, C., Wang, L., 2010. Improving the housing-unit method for small-area population estimation using remote-sensing and GIS information. Int. J. Remote Sens. 31 (21), 5673–5688.
Du, M., Wang, L., Zou, S., Shi, C., 2018. Modeling the census tract level housing vacancy rate with the Jilin1-03 satellite and other geospatial data. Remote Sensing 10 (12), 1920.
Elwood, S., Goodchild, M.F., Sui, D.Z., 2012. Researching volunteered geographic information: Spatial data, geographic research, and new social practice. Ann. Assoc. Am. Geogr. 102 (3), 571–590.
Frazier, A.E., Bagchi-Sen, S., Knight, J., 2013. The spatio-temporal impacts of demolition land use policy and crime in a shrinking city. Appl. Geogr. 41, 55–64.
Griffin, T.L., Yang, E., Flournoy, M., Bartocci, J., 2015. Mapping America’s Legacy Cities. The J. Max Bond Center at the Bernard and Anne Spitzer School of Architecture at the City College of New York.
He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778.
Hillier, A.E., Culhane, D.P., Smith, T.E., Tomlin, C.D., 2003. Predicting housing abandonment with the Philadelphia neighborhood information system. J. Urban Affairs 25 (1), 91–106.
Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T., 2014. Caffe: Convolutional architecture for fast feature embedding. In: Proceedings of the 22nd ACM international conference on Multimedia, pp. 675–678.

Kang, J., Körner, M., Wang, Y., Taubenböck, H., Zhu, X.X., 2018. Building instance classification using street view images. ISPRS J. Photogramm. Remote Sens. 145, 44–59.
Koch, D., Despotovic, M., Sakeena, M., Döller, M., Zeppelzauer, M., 2018. Visual estimation of building condition with patch-level convnets. In: Proceedings of the 2018 ACM Workshop on Multimedia for Real Estate Tech. ACM, pp. 12–17.
Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems, pp. 1097–1105.
Kumagai, K., Matsuda, Y., Ono, Y., 2016. Estimation of housing vacancy distributions: basic Bayesian approach using utility data. Int. Archiv. Photogramm. Remote Sens. Spatial Inf. Sci. 41, 709.
Kurth, J., MacDonald, C., 2015. Volume of Abandoned Homes 'Absolutely Terrifying'. The Detroit News.
Laumer, D., Lang, N., van Doorn, N., Mac Aodha, O., Perona, P., Wegner, J.D., 2020. Geocoding of trees from street addresses and street-level images. ISPRS J. Photogramm. Remote Sens. 162, 125–136.
Law, S., Paige, B., Russell, C., 2019. Take a look around: using street view and satellite images to estimate house prices. ACM Trans. Intell. Syst. Technol. (TIST) 10 (5), 1–19.
Li, X., Zhang, C., Li, W., Ricard, R., Meng, Q., Zhang, W., 2015. Assessing street-level urban greenery using Google Street View and a modified green view index. Urban For. Urban Greening 14 (3), 675–685.
Li, X., Zhang, C., Li, W., 2017. Building block level urban land-use information retrieval based on Google Street View images. GIScience Remote Sens. 54 (6), 819–835.
Li, Z., Long, Y., 2019. Analysis of the variation in quality of street space in shrinking cities based on dynamic street view picture recognition: A case study of Qiqihar. In: Shrinking Cities in China. Springer, Singapore, pp. 141–155.
Lin, C.Y., Chiu, Y.C., Ng, H.F., Shih, T.K., Lin, K.H., 2020. Global-and-local context network for semantic segmentation of street view images. Sensors 20 (10), 2907.
Lynch, A.J., Mosbah, S.M., 2017. Improving local measures of sustainability: A study of built-environment indicators in the United States. Cities 60, 301–313.
Mallach, A., 2012. Depopulation, market collapse and property abandonment: Surplus land and buildings in legacy cities. Rebuilding America's legacy cities: New directions for the industrial heartland, pp. 85–110.
Mallach, A., 2018. The Empty House Next Door. Lincoln Institute of Land Policy, Cambridge, MA.
Molloy, R., 2016. Long-term vacant housing in the United States. Reg. Sci. Urban Econ. 59, 118–129.
Morckel, V.C., 2013. Empty neighborhoods: Using constructs to predict the probability of housing abandonment. Hous. Policy Debate 3 (3), 469–496.
Morckel, V.C., 2014a. Predicting abandoned housing: does the operational definition of abandonment matter? Commun. Dev. 45 (2), 122–134.
Morckel, V.C., 2014b. Spatial characteristics of housing abandonment. Appl. Geogr. 48, 8–16.
Raleigh, E., Galster, G., 2015. Neighborhood disinvestment, abandonment, and crime dynamics. J. Urban Affairs 37 (4), 367–396.
Raman, A., 2019. Cheers to Street View's 10th birthday.
Scafidi, B.P., Schill, M.H., Wachter, S.M., Culhane, D.P., 1998. An economic analysis of housing abandonment. J. Hous. Econ. 7 (4), 287–303.
Silverman, R.M., Yin, L., Patterson, K.L., 2013. Dawn of the dead city: An exploratory analysis of vacant addresses in Buffalo, NY 2008–2010. J. Urban Affairs 35 (2), 131–152.
Simonyan, K. and Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.
U.S. Government Accountability Office, 1978. Housing Abandonment: A National Problem Needing New Approaches. U.S. Government Accountability Office, Washington, DC.
U.S. Government Accountability Office, 2011. Vacant Properties: Growing Number Increases Communities' Costs and Challenges. U.S. Government Accountability Office, Washington, DC.
Wang, L., Fan, H., Wang, Y., 2019. An estimation of housing vacancy rate using NPP-VIIRS night-time light data and OpenStreetMap data. Int. J. Remote Sens. 1–23.
Weiss, K., Khoshgoftaar, T.M., Wang, D., 2016. A survey of transfer learning. J. Big Data 3 (1), 9.
Wiechmann, T., Pallagst, K.M., 2012. Urban shrinkage in Germany and the USA: A comparison of transformation patterns and local strategies. Int. J. Urban Reg. Res. 36 (2), 261–280.
Yao, Y., Li, Y., 2011. House vacancy at urban areas in China with nocturnal light data of DMSP-OLS. In: Proceedings 2011 IEEE International Conference on Spatial Data Mining and Geographical Knowledge Services. IEEE, pp. 457–462.
Yin, L., Silverman, R., 2015. Housing abandonment and demolition: Exploring the use of micro-level and multi-year models. ISPRS Int. J. Geo-Inf. 4 (3), 1184–1200.
Zeppelzauer, M., Despotovic, M., Sakeena, M., Koch, D., Döller, M., 2018. Automatic prediction of building age from photographs. In: Proceedings of the 2018 ACM on International Conference on Multimedia Retrieval. ACM, pp. 126–134.
Zhang, F., Wu, L., Zhu, D., Liu, Y., 2019. Social sensing from street-level imagery: A case study in learning spatio-temporal urban mobility patterns. ISPRS J. Photogramm. Remote Sens. 153, 48–58.
Zhang, W., Li, W., Zhang, C., Hanink, D.M., Li, X., Wang, W., 2017. Parcel-based urban land use classification in megacity using airborne LiDAR, high resolution orthoimagery, and Google Street View. Comput. Environ. Urban Syst. 64, 215–228.
Zhou, B., Lapedriza, A., Khosla, A., Oliva, A., Torralba, A., 2017. Places: A 10 Million Image Database for Scene Recognition. IEEE transactions on pattern analysis and machine intelligence 40 (6), 1452–1464. https://doi.org/10.1109/TPAMI.2017.2723009.
Zou, S., Wang, L., 2020. Individual vacant house detection in very-high-resolution remote sensing images. Ann. Am. Assoc. Geogr. 110 (2), 449–461.