*geomatics*

MDPI

*Article*

# Building Footprint Identification Using Remotely Sensed Images: A Compressed Sensing-Based Approach to Support Map Updating

**Rizwan Ahmed Ansari [1,*], Rakesh Malhotra [1] and Mohammed Zakariya Ansari [2]**

[1] Department of Environmental, Earth and Geospatial Sciences, North Carolina Central University, Durham, NC 27707, USA; rmalhotra@nccu.edu
[2] Independent Researcher, Pune 411048, India; mails2zakariya@gmail.com
* Correspondence: ransari@nccu.edu

**Abstract:** Semantic segmentation of remotely sensed images for building footprint recognition has been extensively researched, and several supervised and unsupervised approaches have been presented and adopted. The capacity to do real-time mapping and precise segmentation on a significant scale while considering the intrinsic diversity of the urban landscape in remotely sensed data has significant consequences. This study presents a novel approach for delineating building footprints by utilizing the compressed sensing and radial basis function technique. At the feature extraction stage, a small set of random features of the built-up areas is extracted from local image windows. The random features are used to train a radial basis neural network to perform building classification; thus, learning and classification are carried out in the compressed sensing domain. By virtue of its ability to represent characteristics in a reduced dimensional space, the scheme shows promise in being robust in the face of variability inherent in urban remotely sensed images. Through a comparison of the proposed method with numerous state-of-the-art approaches utilizing remotely sensed data of different spatial resolutions and building clutter, we establish its robustness and prove its viability. Accuracy assessment is performed for segmented footprints, and comparative analysis is carried out in terms of intersection over union, overall accuracy, precision, recall, and F1 score. The proposed method achieved scores of 93% in overall accuracy, 90.4% in intersection over union, and 91.1% in F1 score, even when dealing with drastically different image features. The results demonstrate that the proposed methodology yields substantial enhancements in classification accuracy and decreases in feature dimensionality.

**Keywords:** urban analysis; building identification; texture classification; remote sensing; compressed sensing

check for updates

Academic Editor: Enrico Corrado Borgogno Mondino

Received: 22 December 2024
Revised: 29 January 2025
Accepted: 29 January 2025
Published: 31 January 2025

**Citation:** Ansari, R.A.; Malhotra, R.; Ansari, M.Z. Building Footprint Identification Using Remotely Sensed Images: A Compressed Sensing-Based Approach to Support Map Updating. *Geomatics* **2025**, *5*, 7. https://doi.org/10.3390/geomatics5010007

**Copyright:** © 2025 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

The swift pace of urbanization and escalating population influx into metropolitan areas necessitate the development of digital and physical infrastructure to meet the demands of new entrants and the densification of urban areas. Consequently, it is important to plan and control the utilization of land in metropolitan areas. Building detection plays a crucial role in urban planning and analysis due to its significant impact on the design and control of space. The process of identifying and mapping built-up regions is crucial for several applications such as urban analysis, geoinformation processing, map updates, change assessment, disaster management, and transportation planning.

Identifying and delineating built-up regions is challenging due to the varied characteristics of these categories, which necessitate the extraction and analysis of spatial information and the consideration of topological distribution with scale difficulties [1]. Although there are automated processes involved in this activity, there is a significant reliance on manual annotations. This manual process is time-consuming and susceptible to errors due to the existence of observational variances. These variances may arise from occlusions and shadows, or from the presence of spectrally similar objects, such as roads, small open areas, and parking lots. Therefore, there is a need for expeditious and dependable automated segmentation techniques. Conventional approaches employ texture, form, spectral, and geographic characteristics, which are then processed using clustering or classification algorithms [2,3].

Texture is present everywhere in natural images and plays a crucial role as a visual clue in several image analysis tasks such as image segmentation, image retrieval, and shape from texture. Texture classification is a crucial concern in the fields of computer vision and image processing since it has a substantial impact on various applications such as medical image analysis, remote sensing, object recognition, and content-based image retrieval [4].

A texture classification system primarily consists of the following two main stages: (1) extracting features and (2) performing classification [5]. The body of research on extracting texture features is significant, with comprehensive assessments [6–8]. Prominent approaches commonly used include Gray Level Cooccurrence Histograms [9], Markov Random Fields [10,11], Local Binary Patterns [12,13], Gabor filters [14,15], textures from elevation maps [16,17], curvelet and contourlet-based multiresolution techniques [18,19], wavelet textures [20–22], and fractal techniques [23–25]. Second-order statistics and their inter-relationships from gray level co-occurrence matrices have been used to represent different smooth and rough textures for classification [9]. A model based on a three-dimensional Gaussian Markov Random field is proposed in [11] for volumetric texture segmentation. This method adaptively utilizes the texture cubes, followed by k-means clustering to improve the classification accuracy. A scale- and pattern-adaptive local binary pattern algorithm is proposed to minimize scale sensitivity and the effects of noisy rotational characteristics in texture classification. Kirsch gradient operators were used to extract multiscale features for improved accuracy [12]. Gabor filters have been used to effectively represent multi-frequency and detailed localized information. This method improved the accuracy for fine-grained texture [15]. Texture features using elevation maps are fused with the point clouds to extract building footprints. Texture features were used to represent the height of each point in the cloud [16]. Multiresolution techniques were used to capture different scales of texture for remote sensing image classification. Curvelet- and contourlet-based textures improved the accuracy over wavelet-based textures [18,19]. These methods select a small number of texture features from local picture patches that are often fewer than the dimensionality of the original image patch. Most feature extraction methods primarily concentrate on local texture information by analyzing the gray level patterns surrounding a specific pixel. However, texture is also defined by its overall appearance, which includes the repetition of local patterns and the interconnections between them.

The effectiveness of the sparse technique heavily relies on the texture images, some of which may not provide a sufficient number of regions to create a strong and reliable representation of the texture. Consequently, the dense method is more prevalent and extensively researched [12].

The crucial factor in patch-based classification is the window size. Small window sizes are inadequate for capturing large-scale structures that may be the primary characteristics of certain textures. They are also not resilient to local changes in textures and are particularly susceptible to noise and missing pixel values resulting from fluctuations in

lighting [11,12]. Nevertheless, patch representation has a drawback in that the dimension of the patch space increases quadratically with the size of the window. This high dimensionality presents two issues for the classification techniques employed in texton learning. The presence of irrelevant and noisy features can potentially mislead the classification algorithm. Furthermore, in high-dimensional spaces, data often exhibits sparsity, which poses challenges in accurately representing the underlying structure of the data [11].

Thus, it is reasonable to investigate if patch vectors in high dimensions may be transformed into a lower-dimensional subspace without experiencing significant loss of information. A low-dimensional feature space has several advantages, including less storage needs, decreased computing complexity, and the ability to overcome the curse of dimensionality, resulting in improved classification performance. Using a limited yet noticeable collection of features would make it easier to represent patterns and classify them. However, commonly used strategies to reduce the number of dimensions sometimes lead to a loss of information during the projection process. This introduces us to the field of compressive sensing.

The compressed sensing (CS) approach, which has been the inspiration for this research, is attractive due to the remarkable finding that high-dimensional sparse data may be precisely reconstructed using only a small number of nonadaptive linear random projections. When utilizing CS in texture classification, the main concern is how effectively random projections can retain information about high-dimensional sparse texture signals in local image patches and whether this retention provides any benefits in classification.

The theory of compressed sensing has gained prominence due to the research conducted by the authors in [26,27]. They have demonstrated the benefit of using random projections to capture information about signals that are sparse or compressible. The fundamental concept of compressive sensing is based on the idea that a limited set of nonadaptive linear measurements of a compressible signal or picture can provide sufficient information to achieve almost flawless reconstruction and processing. This evolving idea has sparked studies in various applied fields, including image reconstruction [28,29], vehicular networks [30], medical image classification and analysis [31–33], underwater image analysis [34], hyperspectral image compression [35,36], remote sensing applications [37–39], and machine learning [40–43], among others.

The efficacy of CS in order to achieve accurate signal reconstruction has been demonstrated in previous studies [26,27,44–48]. The authors in [45] developed a real-time technique for detecting and localizing flaws utilizing Gaussian mixture-based local gray-scale patches for texture characterization, followed by classification through a multiscale framework, which demonstrated enhanced accuracy compared to non-compressed methods. A multi-layer basis pursuit framework is proposed in [46], combining the benefits of objective-based compressed sensing reconstructions and deep learning-based methods via iterative thresholding algorithms to efficiently train compressed sensing MRI image restoration on GPU, resulting in accelerated convergence and improved peak signal-to-noise ratio. A technique utilizing compressed sensing and the k-nearest neighbor classifier has been developed to improve lung cancer detection for decision support in clinical diagnosis [47].

However, the utilization of CS in the context of texture classification for the purpose of building identification has received limited attention. Previous studies [49–52] have focused on utilizing the unique structure of sparse coding for texture patches. The discrete cosine transform for sparsification is employed in patches with the CNN AlexNet for vegetable classification with orthogonal matching pursuit [50]. Block-based distributed compressed sensing was developed to leverage both intra- and inter-correlation structures of hyperspectral images, facilitating a high compression ratio at low sampling rates [51].

These studies have explored the recovery process and have carefully designed a sparsifying redundant dictionary. However, this research carries out classification in the compressed domain without depending on any reconstruction. This study showcases an experiment that demonstrates the advantages of this innovative theory for texture categorization. The proposed method is characterized by its computational simplicity while still possessing significant efficacy. Instead of conducting texture classification in the original high-dimensional patch space or attempting to determine an appropriate feature extraction approach, we employ random projections of local patches to carry out texture classification in a significantly lower-dimensional compressed patch space. According to the theory of CS, the specific selection of the number of features is no longer crucial. As long as the number of random features exceeds a certain threshold, it will include sufficient information to maintain the underlying local texture structure and accurately categorize the given test image.

## 2. Materials and Methods

### 2.1. Datasets

The proposed methodology is tested using three data sets. A synthetic test image from a mosaic of four Brodatz texture images [53] is created as a test bed. This test bed is used as a proof of concept to extract feature descriptors in the compressed sensing domain, as the input consists of pure texture samples. In the absence of a target class, binary classification is not conducted; therefore, class-wise accuracies are computed. Each texture image is of size $256 \times 256$ with 256 gray levels, which yields the test bed image of size $512 \times 512$.

In this study, we assessed the effectiveness of the proposed method by testing it on the WHU dataset [54] and the OpenCities dataset [55]. The WHU data collection consists of around 220,000 distinct buildings that were taken from aerial photos with a spatial resolution of 7.5cm. These structures occupy an area of 450 $km^2$ and represent the city of Christchurch in New Zealand. The OpenCities collection comprises 790,000 building footprints derived from OpenStreetMap, representing 10 cities and regions in Africa. The aerial imaging resolution varies across different regions, ranging from 0.02m to 0.2m. This dataset comprises the images and the building footprints contained in the GeoJSON files. All training images have been reprojected to the corresponding UTM zone projection for their respective regions. The building ratio, defined as the proportion of area occupied by structures, varies from 0.19 to 0.36, with an average building size ranging from 47.43 $m^2$ to 150.71 $m^2$ [55]. The studies utilized aerial photography from different cities to create a comprehensive dataset that includes both highly and sparsely populated locations. The combined dataset covers a wide range of regions, including rural landscapes and industrial zones with various architectural styles and density. For the experiments, we randomly partition each dataset into training, validation, and test sets at a ratio of 7:2:1.

### 2.2. Compressed Sensing

Compressed sensing takes advantage of the fact that many types of signals have a structure that is lower in dimensionality relative to the larger space in which they exist. Compressed sensing states that, for some signal types, a limited number of nonadaptive measurements in the form of randomized projections can effectively capture most of the important information in a signal and provide a good approximation of the original signal. The advantage of CS theory lies in its ability to perfectly recover a signal that can be represented sparsely on a set of basis functions using a very small number of random projections.

The fundamental principle of CS is based on the concept of signal sparsity or compressibility, and the compressibility of textures is well recognized and accepted. Undoubtedly,

through significant experience, the wavelet transform has proven that most natural images can be compressed [56,57]. Textures, due to their fixed or periodic nature, tend to be less abundant. In addition, the extensive body of literature on texture classification based on extracting features from small image patches reveals that textures have a limited number of degrees of freedom. The author in [58] employs a filter bank to initially decrease the patch space. Subsequently, the dimension of texture filter responses is further reduced by projecting filter marginals onto a low-dimensional manifold. This demonstrates that by projecting onto a manifold of an appropriate dimension, the accuracy of classification can be enhanced.

In CS, a fundamental assumption is that of sparsity or compressibility. Let $\underline{y} \in \mathbb{R}^{n\times 1}$ be an unknown signal of length $n$, and let $\psi = [\psi_1, \psi_2, \ldots \psi_n]$ be an orthonormal basis where $\underline{\psi}_i \in \mathbb{R}^{n\times 1}$, such that as in [27].

$$\underline{y} = \sum_{i=1}^{n} \theta_i \underline{\psi}_i = \psi\underline{\theta}, \tag{1}$$

where $\underline{\theta} = [\theta_1, \theta_2, \ldots, \theta_n]$ denotes the vector of coefficients that represents $\underline{y}$ in the basis $\psi$. The signal $\underline{y}$ is considered sparse if the majority of the coefficients in $\underline{\theta}$ are zero or may be deleted without significant loss of information.

Let $\phi$ be an $m \times n$ sampling matrix, with $m << n$ such that as in [27].

$$\underline{x} = \phi\underline{y} = \phi\psi\underline{\theta}, \tag{2}$$

where $\underline{x}$ is an $m \times 1$ vector of linear measurements.

The sampling matrix $\phi$ must allow the reconstruction of a length-$n$ signal $\underline{y}$ from a length-$m$ measurement vector $\underline{x}$. Generally, there is a loss of information while using this transformation as it exhibits dimensionality reduction. However, it has been demonstrated that the measurement matrix can preserve the information in sparse and compressible signals, provided it satisfies the restricted isometry property.

The signal reconstruction algorithm must utilize the m data, the random measurement in $\underline{x}$, and the basis $\psi$ in order to recreate $\underline{\theta}$.

### *2.3. Feature Extraction*

There are C unique texture categories, each containing S instances. Consider an ensemble $\{I_{C,S}\}_{s=1}^{S}$ to represent the samples of class c, and let the full texture collection be denoted as $D = \left\{\{I_{C,S}\}_{s=1}^{S}\right\}_{c=1}^{C}$. A set of $\sqrt{n} \times \sqrt{n}$ image patches $P$ is extracted from image $I_{c;s}$ and the compressed sensing measures $\underline{x} = \phi\underline{p}$ obtained from features $\underline{p}$, are employed. In this study, we opt to use a Gaussian random matrix $\phi$, which means that the entries of the matrix are independent and follow a normal distribution with zero mean and unit variance.

The compressed domain refers to the state or representation of data after it has been compressed [27].

$$\chi = \left\{\underline{x} = \phi\underline{p} \mid \underline{p} \in \mathrm{P}\right\}, \tag{3}$$

Therefore, it is a compressed depiction of the patch domain as in [27].

$$\mathrm{P} = \left\{\underline{p} \mid \underline{p} \in \mathbb{R}^{n\times 1}\right\} \tag{4}$$

The classification stage consists of the following steps:

1. The compressed texton dictionary learning stage involves directly learning a universal compressed texton dictionary W in the compressed domain X, rather than from the patch domain P.

2. In order to identify a buildup area from the input image, a machine learning model that has been trained for semantic segmentation is applied to the texton vector that contains the extracted features in step (1). Our classifier system employs a neural network that is based on the radial basis function as its learning paradigm.

Figure 1 illustrates the proposed workflow through training and testing phases.

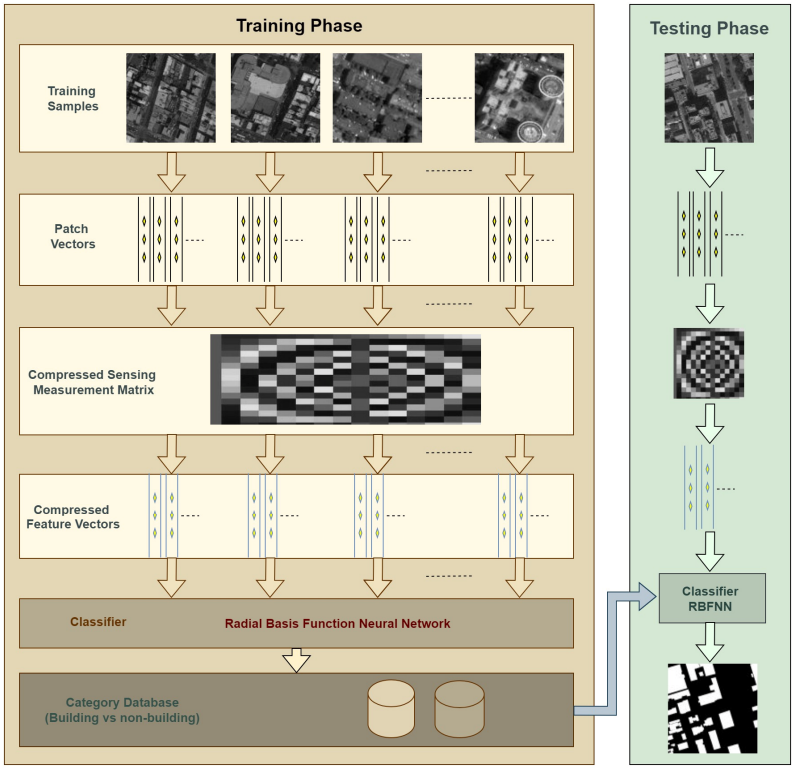


**Figure 1.** Proposed methodology.

### *2.4. Radial Basis Function Classifier*

The radial basis function (RBF) network is a three-layer feedforward neural network consisting of input, hidden, and output layers, as shown in Figure 2. Every node in the hidden layer employs a radial basis function as its nonlinear activation function. The hidden layer executes a nonlinear transformation of the input. The output layer functions as a linear combiner, transforming the nonlinearity into a different space.

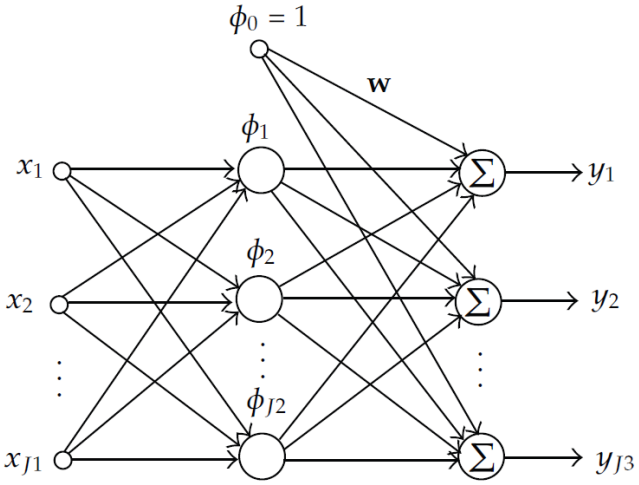


**Figure 2.** Radial basis network (adapted from [59]).

The RBF network utilizing a localized RBF, such as the Gaussian RBF network, is classified as a receptive-field or localized network. The localized approximation method yields the most robust output when the input is proximate to a node's prototype. A properly trained localized RBF network yields identical outputs for proximate input vectors, whereas far input vectors result in almost independent outputs. This is the intrinsic local generalization characteristics. A receptive-field network is an associative neural network wherein only a limited subspace is defined by the network's input. This trait is notably appealing as the alteration of the receptive-field function generates a localized effect. Receptive-field networks can be efficiently created by modifying the parameters of the receptive-field functions and/or by adding or removing neurons. Consequently, we employed an RBF network for the classification phase to encapsulate local texture characteristics within the image patches.

The radial basis function-based neural network (RBFNN) is trained using a supervised learning approach, in which the model is trained on a labelled dataset. Ideally, for prediction, these classes are presumed to be a superset of all the building classes that the model is expected to encounter in the future. The model's generalization capability is subsequently evaluated by comparing its performance on unseen images that comprise a test set.

The RBFNN is a type of artificial neural network that employs the radial basis functions (RBF) for activation [59]. A neuron's output is expressed as a linear combination of the radial basis function of its inputs and parameters. Within our CS system, the layers under consideration consist of 67, 600, and 32 artificial neurons, respectively. The activations of the input layer to the hidden layer are controlled by a Gaussian kernel specified with a spread of 2. The use of a kernel offers the benefit of converting the features into a higher-dimensional space, hence enabling their linear separability. The activation functions from the hidden layer to the output layer are determined by a set of weights that are tuned using a weighted mean squared error-based cost minimization technique. Algorithm 1 shows the training steps in the RBF network.

Compressed Sensing-Radial Basis Function Algorithm

**Algorithm 1**. Training the CS-RBF model

**Inputs:** The original images and texture features.
The cross-entropy error is used as the loss function.
**Output:** Weight and bias matrices; the predicted output of the CS-RBF (label values)
**Procedure:**
1: Initialize learning rate, batch size, kernel size, number of kernels, number of max iterations, dropout, and so on.
2: Generate random weights with a Gaussian type and biases with 0;
CS-RBF_model = InitCS-RBF_model (weights and bias matrices);
3: **While** iter < max iteration or error > min error **do**
Compute error according to loss function
**For** iter = 1 to iter < = number/(batch size) **do**
CS-RBF_model.train (TrianingData and TraingLabels), as loss is minimized with gradient descent; update weight and bias matrices;
**end for**
iter ++
**end while**
4: Save parameters (weight, bias) of the CS-RBF;
5: Training CS-RBF finished.

### *2.5. Performance Metrics and Comparison with Other Methods*

The predominant metrics employed to assess a binary classification (building vs. non-building) technique are precision and recall. The precision is computed as the ratio of predicted building pixels classified as buildings, while recall is calculated as the ratio of all labeled building pixels that are correctly predicted as defined in Equations (5)–(6) [60]. F-score, overall accuracy (OA), and intersection over union (IoU) are utilized for quantitative evaluation as defined in Equations (7)–(9) [60].

- TP (true positive): represents the number of building pixels that have been properly classified as buildings.
- FP (false positive): represents the number of non-building pixels being misclassified as buildings.
- FN (false negative): represents the number of building pixels being misclassified as non-buildings.
- TN (true negative): represents the number of non-building pixels that have been properly classified as non-buildings.

The performance metrics are as follows, as defined in [60]:

$$\text{Precision} = \text{TP}/(\text{TP} + \text{FP}) \tag{5}$$

$$\text{Recall} = \text{TP}/(\text{TP} + \text{FN}) \tag{6}$$

$$\text{F-score} = 2\text{TP}/(2\text{TP} + \text{FP} + \text{FN}) \tag{7}$$

$$\text{Overall accuracy} = (\text{TP} + \text{TN})/(\text{TP} + \text{TN} + \text{FN} + \text{FP}) \tag{8}$$

$$\text{IoU} = \text{TP}/(\text{TP} + \text{FP} + \text{FN}) \tag{9}$$

Various state-of-the-art methods have been implemented and employed for comparative analysis with the proposed method. The list comprises U-Net [61], Deep ResUnet [62], ScattNet [63], UNetFormer [64], CG-Swin [65], ASF-Net [66], ResiDualGAN [67], and SA-MRA [68]. The U-Net architecture [61] is a streamlined network that omits fully connected layers and exclusively employs the valid segment of each convolution. Furthermore, it utilizes comprehensive data augmentation by implementing elastic deformations on the training images. This enables the network to attain invariance to these deformations, irrespective of the presence of such alterations in the labeled set. To introduce ResUnet, the authors in [62] replaced conventional neural units with residual units as core elements of the U-Net architecture. This network outperforms U-Net while utilizing only 25% of its parameters.

All models were developed utilizing the TensorFlow platform (version 1.13) and trained on the same dataset according to the intended network until they exhibited signs of overfitting.

## 3. Results

Figure 3 illustrates the Brodatz mosaic synthetic image along with its associated segmentation result. As the true classification for this test bed mosaic with four classes is known, class-wise accuracy and confusion matrix along with the kappa coefficient are computed to demonstrate the ability of the proposed method. Table 1 details the accuracy analysis of the segmentation results. The proposed method produced a very good separation with minor errors near the texture boundaries, yielding a Kappa coefficient of 0.937.

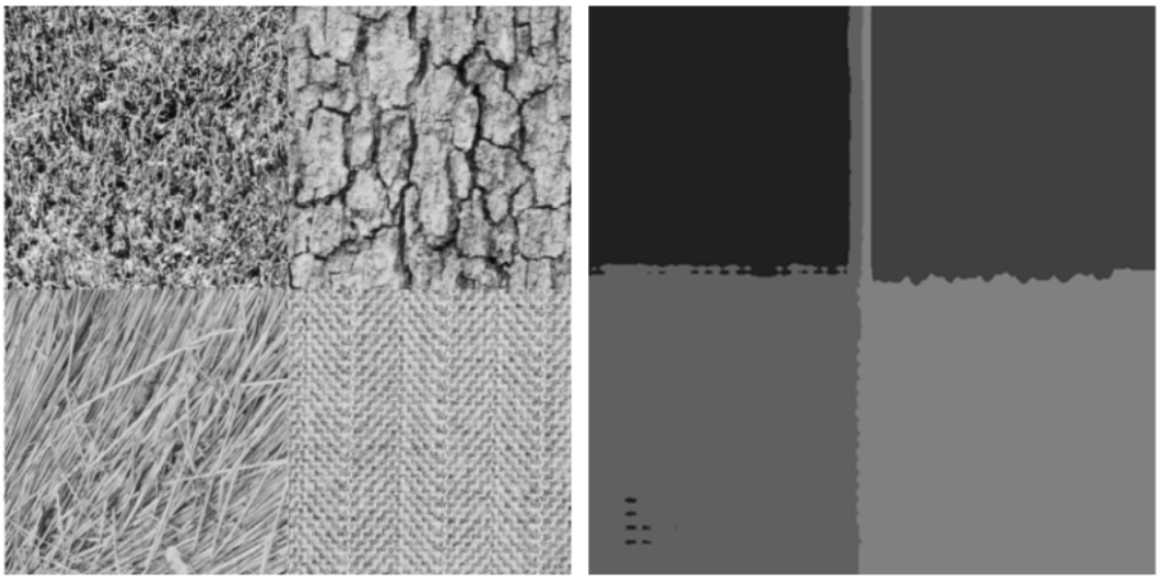


**Figure 3.** Segmentation for synthetic test bed image: (**a**) original image compiled from Brodatz set [53] and (**b**) segmented image using the proposed scheme.

**Table 1.** Accuracy assessment for the Brodatz test bed image.

| | **Class 1 (%)** | **Class 2 (%)** | **Class 3 (%)** | **Class 4 (%)** |
|---|---|---|---|---|
| Class 1 (%) | 90.34 | 6.28 | 0.36 | 3.01 |
| Class 2 (%) | 0 | 95.8 | 0 | 4.19 |
| Class 3 (%) | 0.25 | 0 | 94.85 | 4.89 |
| Class 4 (%) | 0 | 0 | 0 | 100 |
| Kappa coefficient = 0.937 | | | Kappa error = 0.0006 | |

Figure 4 displays the results for the WHU dataset. It presents four samples from the WHU dataset, exhibiting varying built-up area densities, different structural characteristics, and both random and recurring patterns of buildings and roadways. The first two columns display the input image samples along with their respective ground truth images. The first row in Figure 4 shows the enhanced segmentation results produced by the proposed method, effectively delineating unique and actual building borders, unlike the U-Net, which yields arbitrary oblong shapes as building footprints, as seen in Figure 4c with an accuracy of 89.70%. The second row of Figure 4 c,e shows how the UNet and ResiDualGAN methodologies often merge closely situated buildings into a single building block. The ResiDualGAN approach (Figure 4e) classifies non-built-up regions as buildings in the third row, yielding an overall accuracy of 89.80%. A non-building area is classified as a building class by UNet, as evidenced in the last row of Figure 4c. The semantic segmentation results demonstrate that the proposed scheme outperforms other conventional and deep learning frameworks in terms of segmentation quality. Our approach demonstrates a 3.5% improvement in mean Intersection over Union (mIoU) and a 3.7% improvement in total accuracy compared to previously used methods (Table 2).

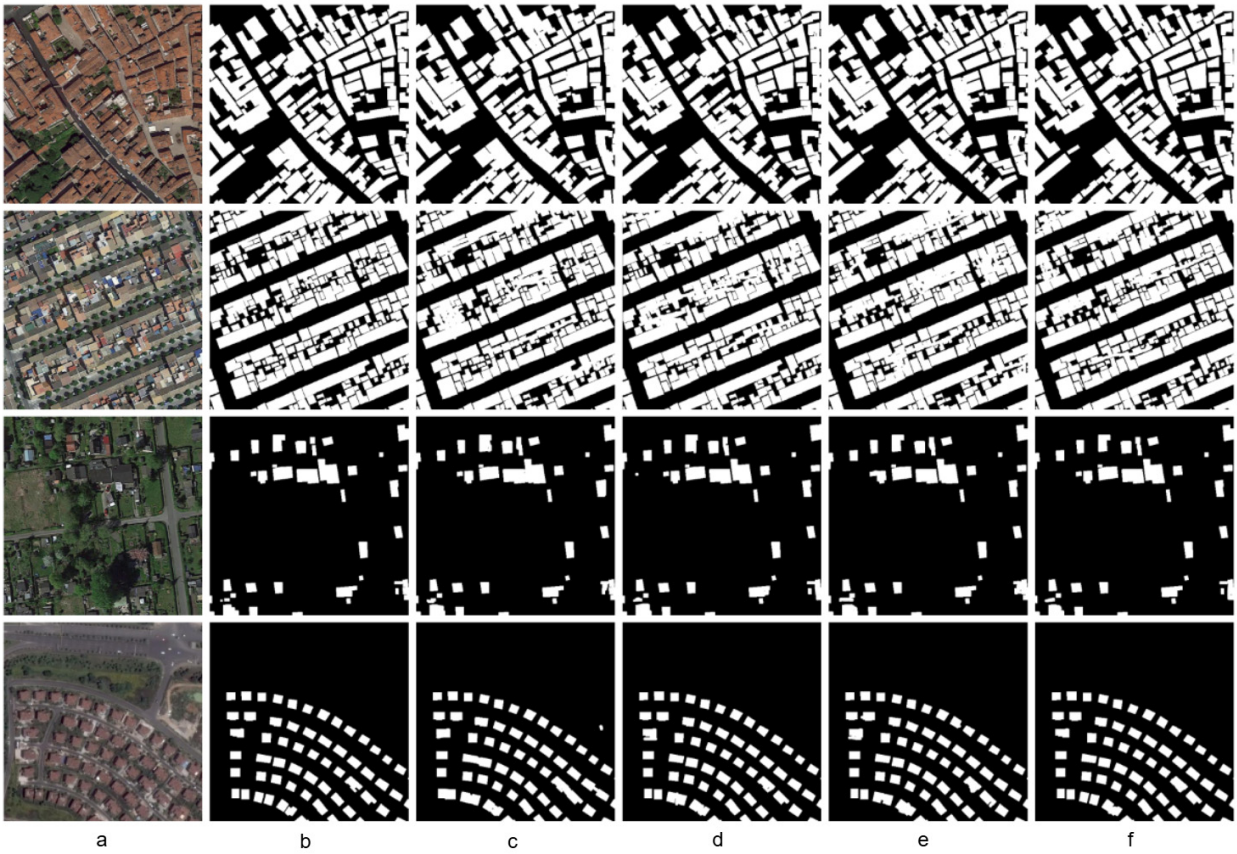


**Figure 4.** Column-wise: (**a**) input image, (**b**) ground truth, and segmentation results of (**c**) UNet, (**d**) ResUNet, (**e**) ResiDualGAN, and (**f**) proposed method on the WHU building dataset with four sample images (row-wise).

**Table 2.** Performance comparison for the WHU dataset.

| Model | IoU | | mIoU | Precision | Recall | F-Score | Accuracy (%) |
|---|---|---|---|---|---|---|---|
| | Building | Non-Building | | | | | |
| U-Net [61] | 0.876 | 0.906 | 0.891 | 0.853 | 0.901 | 0.876 | 89.70 |
| Deep ResUnet [62] | 0.719 | 0.854 | 0.786 | 0.901 | 0.875 | 0.887 | 90.30 |
| ScattNet [63] | 0.904 | 0.918 | 0.911 | 0.882 | 0.821 | 0.850 | 89.03 |
| UNetFormer [64] | 0.903 | 0.859 | 0.881 | 0.883 | 0.892 | 0.887 | 92.96 |
| CG-Swin [65] | 0.917 | 0.889 | 0.903 | 0.902 | 0.885 | 0.893 | 90.61 |
| ASF-Net [66] | 0.892 | 0.901 | 0.896 | 0.886 | 0.902 | 0.893 | 91.89 |
| ResiDualGAN [67] | 0.793 | 0.842 | 0.817 | 0.891 | 0.876 | 0.883 | 89.80 |
| SA-MRA [68] | 0.898 | 0.936 | 0.917 | 0.887 | 0.902 | 0.894 | 93.24 |
| Proposed method | 0.904 | 0.948 | 0.926 | 0.904 | 0.919 | 0.911 | 93.41 |

To further demonstrate the reliability of the proposed approach, the outcomes have been compared with the OpenCities dataset. In Figure 5, the qualitative results are presented. The first row displays aerial photographs captured over Accra, while the second row displays aerial shots of Kampala. In the Accra image, the proposed approach can accurately identify clean and distinct building footprints, even in areas with significant clutter, as demonstrated in the subsequent cropping. In the second row (Figure 5f), the buildup portions are accurately segmented with fine border details. The structural details of built-up areas are retained while correctly segregating roads and open spaces. Non-building pixels are incorrectly segmented as buildings by other methods, as demonstrated in the second row in Figure 5c–e. Similarly, in the fourth row in Figure 5, the proposed method

effectively identifies and extracts narrow building structures while significantly reducing misclassifications evident when compared with the UNet, ResUNet, and ResiDualGAN techniques. Self-attention network [68] performs better among the existing methods with an overall accuracy of 91.24% since it uses a multiresolution-based framework to extract multiscale features of the buildings. However, the proposed approach yields an improved mean Intersection over Union (mIoU) of 0.907 and pixel accuracy of 92.11%. The achieved F-score of 0.894 indicates high accuracy in identifying building regions (Table 3). The results of this study demonstrate the effectiveness of the proposed compressed sensing method with the RBF classifier for extracting building footprints for both the datasets used.

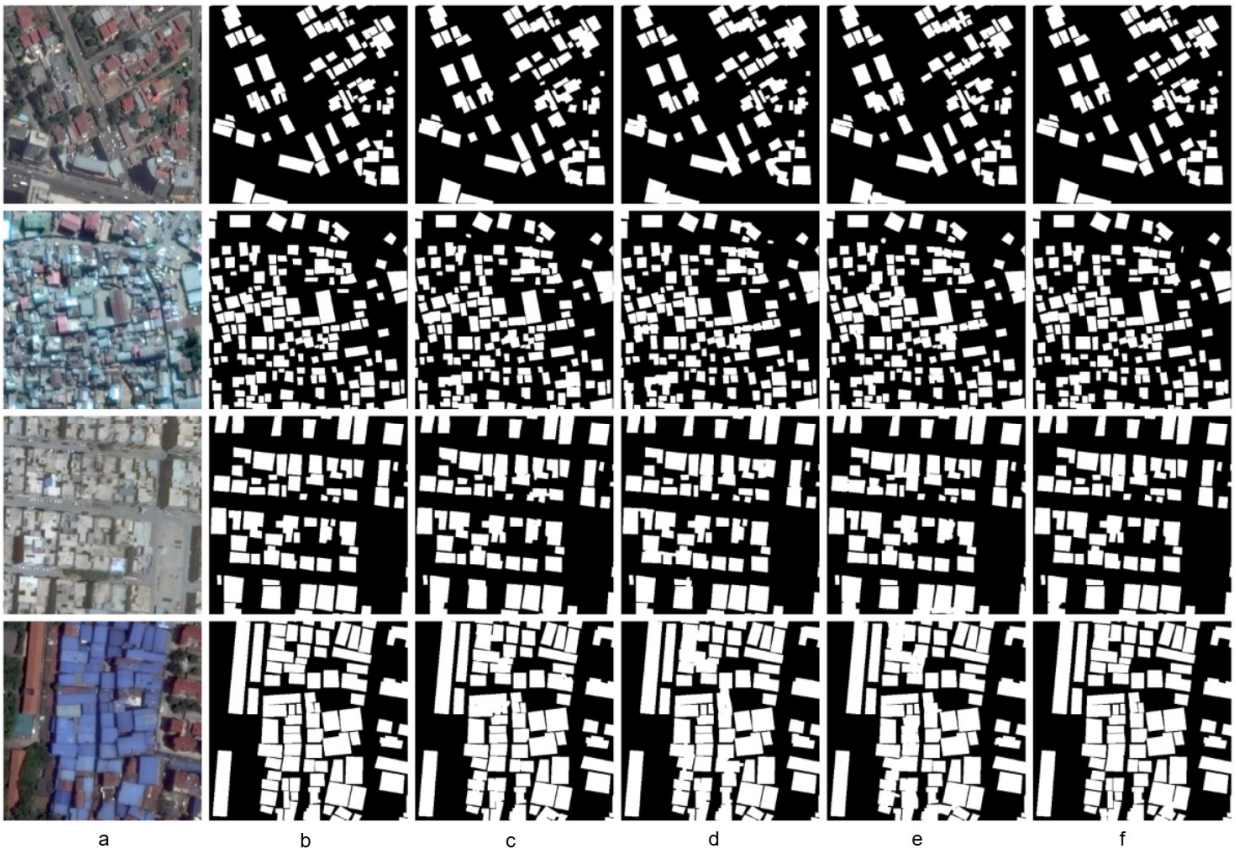


**Figure 5.** Column-wise: (**a**) input image, (**b**) ground truth, and segmentation results of (**c**) UNet, (**d**) ResUNet, (**e**) ResiDualGAN, and (**f**) proposed method on the OpenCities building dataset with four sample images (row-wise).

**Table 3.** Performance comparison for the OpenCities dataset.

| Model | IoU | | mIoU | Precision | Recall | F-Score | Accuracy (%) |
|---|---|---|---|---|---|---|---|
| | Building | Non-Building | | | | | |
| U-Net [61] | 0.882 | 0.896 | 0.889 | 0.842 | 0.895 | 0.867 | 89.21 |
| Deep ResUnet [62] | 0.726 | 0.846 | 0.786 | 0.757 | 0.801 | 0.778 | 82.10 |
| ScattNet [63] | 0.793 | 0.864 | 0.828 | 0.842 | 0.802 | 0.821 | 84.03 |
| UNetFormer [64] | 0.885 | 0.843 | 0.864 | 0.827 | 0.892 | 0.858 | 87.96 |
| CG-Swin [65] | 0.901 | 0.873 | 0.887 | 0.878 | 0.863 | 0.870 | 89.61 |
| ASF-Net [66] | 0.881 | 0.911 | 0.896 | 0.881 | 0.893 | 0.886 | 90.89 |
| ResiDualGAN [67] | 0.784 | 0.832 | 0.808 | 0.829 | 0.858 | 0.843 | 84.80 |
| SA-MRA [68] | 0.901 | 0.895 | 0.898 | 0.895 | 0.868 | 0.881 | 91.24 |
| Proposed method | 0.898 | 0.916 | 0.907 | 0.891 | 0.898 | 0.894 | 92.11 |

## 4. Discussion

The demonstrated improvements in quantitative metrics highlight the benefits and usefulness of a CS framework combined with RBFNN in enhancing the quality of semantic segmentation of remotely sensed pictures for building extraction. The first row in Figure 4 exhibits the enhanced segmentation results produced by the proposed method in its capacity to extract distinct and realistic building geometry masks from the aerial photos, in contrast to a U-Net, which produces random oblong blobs as building footprints. The proposed network not only achieves precise extraction of straight-edged buildings but also provides very high accuracy in extracting curvilinear-shaped buildings in terms of structure. This contrasts with ResUnet, which produces building footprints that are highly noisy and do not have consistent semantics. The trade-off between precision and recall, within the framework of binary semantic segmentation of buildup areas, might be presented as follows. A higher recall indicates a model's improved capacity to properly detect many building pixels, albeit at the expense of misclassifying certain background pixels as buildings, hence diminishing the precision score as illustrated in Table 2. By demonstrating the capability of the CS framework coupled with the RBFNN to accurately extract very thin building structures while preserving the original shape and orientation of the built-up structure, the superiority of the method has been further shown in the second row of Figure 4 (another sample from the WHU dataset).

The U-Net and ResUnet models lack the capability to extract individual footprints and instead group all the building masks into single large masses. This characteristic is not well-suited for many urban planning applications. The findings in [65], which used a category-level semantic information modeling approach to improve class segmentation accuracy in remotely sensed images, are consistent with this.

The experimental results on the OpenCities and WHU datasets indicated that the proposed method exhibited better performance than the comparison methods in identifying building area class. The efficacy of the proposed CS-RBF approach can be attributed to the following factors: first, each ground object in the various image scenes exhibited distinct texture, and the reflected intensity value varied with changing orientation angle and scales. This changing trend could contribute to the identification of the label of different types of objects. This distinct exhibition of textural characteristics also facilitated cancer diagnosis in medical imaging, as reported in [47]. Similarly, the k-texture technique for satellite image segmentation used the spatial data about texture bands to improve classification accuracy [2]. The distinct texture features were utilized in the OreInst network, an efficient segmentation technique that has a lightweight structure due to compressed and more accurate representation of ore features in terms of textures [8]. Second, compared with the deep learning models, the CS approach exploited more prior knowledge about nonlinear structure as distinct extracted features were given to the network. As a result, better representations of nonlinear data were generated in seeking the proper features associated with each image patch. The authors in [36] utilized prior knowledge in terms of multiple residual modules and spatial attention to enhance the compression performance of the backbone network for hyperspectral images. The nonlinear feature representation was employed in multiresolution networks in the textural classification method for improved accuracy of remotely sensed images [18,19]. The distinct features of different classes were efficiently represented by the textures of nonlinear representation, which improved the overall accuracy of the algorithm [19]. Additionally, improved representations of different classes were derived using nonlinear basis functions in a change detection methodology of polarimetric synthetic aperture radar images in remote sensing [20]. Third, instead of the forcible aggregation of different classes for each pixel and label, the incorporation of the compressed sensing features established a bridge that explains the inner relationship in

the learned dictionary for building identification. This aligns with findings in [45], which exploited compressed sensing texture surfaces to detect errors in real-time texture error detection with improved detection rates [45]. The authors in [43] utilized compressed sensing features to design an encryption network using a multicolor space and textures. The compressed sensing features exploited the intercorrelations between color spaces to improve encryption security with a high-quality reconstruction of the images.

In densely urbanized areas (first two samples in Figure 4), buildings are often closely packed together, with little green space between them. This can make detection more challenging due to the potential overlap of building footprints (as observed in the second sample of Figure 4), complex roof structures, and shadows. Dense areas may also involve a variety of building types, adding complexity to the detection process.

In low-density building areas (third and fourth samples in Figure 4), where houses are more spread out with significant space between them (often with gardens, driveways, and roads), detection tends to be more efficient. The larger spaces between buildings make it easier to distinguish individual footprints because there is less likelihood of overlap, as observed in the last row of Figure 4. The presence of less structured and irregular building footprints (as seen in the third row in Figure 4, houses with non-rectangular shapes or compound-style layouts) produces overlapping in footprint extraction. When buildings in an area have uniform architectural styles (fourth sample in Figure 4), detection tends to be more efficient as the detection relies on consistent patterns in shape, size, and layout.

In areas with diverse building types (first sample in Figure 5) (e.g., a mix of high-rise apartments, detached houses, industrial buildings, and commercial structures), detection becomes more complex due to the varied shapes, sizes, and layouts of the buildings. The presence of different roof types (flat, pitched, and dome-shaped) and building footprints (rectangular, circular, and irregular) increases the complexity. In some areas, unconventional building types (second and fourth samples in Figure 5, temporary structures, modular homes, or informal dwellings) may present challenges. These buildings may not follow typical architectural norms or geometric patterns, making it difficult to be detected more accurately.

The efficiency of building footprint detection in the datasets used is sensitive to both the landscape type and the variety of buildings present. The WHU dataset produces higher overall accuracy (93.41% vs. 92.11%) than the OpenCities dataset. Areas with homogenous landscapes and building types are easier to process, while diverse, densely built areas present more challenges.

The superior performance of the CS technique compared to other methods clearly demonstrates that the CS matrix effectively retains important information present in the local patch, as anticipated by CS theory. This implies that conducting classification in the compressed patch space is not a drawback but an advantage that can be exploited for efficient representation. In comparison to the patch approach, CS not only provides superior classification accuracy but also operates at a significantly reduced-dimensional feature space, hence decreasing storage needs and computing time. This technique is effective for binary classes; however, in multiclass problems, learning the dictionary and extracting features poses a significant challenge for RBF networks. The selection of kernels for the activation function in the RBF may be sensitive in the context of different real texture modeling.

## 5. Conclusions

This study presents a classification approach that utilizes a representation of textures as a compact collection of CS measures of local texture patches. Our study has demonstrated that CS measurements of local patches can be efficiently utilized for texture classification

for building identification from remotely sensed images. The proposed approach has also demonstrated the ability to equal or exceed the current leading methods in texture categorization while also achieving substantial savings in both time and storage complexity. Approximately one third of the dimensionality of the original patch space is required to maintain the important information present in the initial local patch. Any additional increase in the number of features only results in slight enhancements in classification performance. However, creating a learning dictionary for a multiclass scenario in the remotely sensed images may pose a challenge. The encouraging findings of this study motivate additional investigation into CS-based binary classification. The use of a more advanced classifier, such as a kernel-based nonlinear SVM, may, in certain cases, yield superior classification performance compared to the RBF network classifier employed in the present study. Moreover, the suggested methodology can be integrated into the signature framework, which is currently under investigation in the texture modeling field and is seen as providing some benefits compared to the distance framework. A potential direction for further research could be to expand the proposed framework to address the object-level classification problem with sorted random projections, which differs from the pixel-level classification problem studied in this paper.

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

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

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

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The raw data used in this article are available from the authors on request.

**Acknowledgments:** The authors would like to express profound gratitude to the reviewers and editors who contributed their time and expertise to ensure the quality of this work.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Li, Z.; Xin, Q.; Sun, Y.; Cao, M. A deep learning-based framework for automated extraction of building footprint polygons from very high-resolution aerial imagery. *Remote Sens.* **2021**, *13*, 3630. [CrossRef]
2. Wagner, F.H.; Dalagnol, R.; Sánchez, A.H.; Hirye, M.C.; Favrichon, S.; Lee, J.H.; Mauceri, S.; Yang, Y.; Saatchi, S. K-textures, a self-supervised hard clustering deep learning algorithm for satellite image segmentation. *Front. Environ. Sci.* **2022**, *10*, 946729. [CrossRef]
3. Reska, D.; Kretowski, M. GPU-accelerated image segmentation based on level sets and multiple texture features. *Multimed. Tools Appl.* **2021**, *80*, 5087–5109. [CrossRef]
4. Alsmadi, M.K. Content-based image retrieval using color, shape and texture descriptors and features. *Arab. J. Sci. Eng.* **2020**, *45*, 3317–3330. [CrossRef]
5. Humeau-Heurtier, A. Color texture analysis: A survey. *IEEE Access* **2022**, *10*, 107993–108003. [CrossRef]
6. Yu, Y.; Wang, C.; Fu, Q.; Kou, R.; Huang, F.; Yang, B.; Gao, M. Techniques and challenges of image segmentation: A review. *Electronics* **2023**, *12*, 1199. [CrossRef]
7. Ranjbarzadeh, R.; Sadeghi, S.; Fadaeian, A.; Jafarzadeh Ghoushchi, S.; Tirkolaee, E.B.; Caputo, A.; Bendechache, M. ETACM: An encoded-texture active contour model for image segmentation with fuzzy boundaries. In *Soft Computing*; Springer: New York, NY, USA, 2023; pp. 1–13.

8. Sun, G.; Huang, D.; Peng, Y.; Cheng, L.; Wu, B.; Zhang, Y. Efficient segmentation with texture in ore images based on box-supervised approach. *Eng. Appl. Artif. Intell.* **2024**, *128*, 107490. [CrossRef]
9. Zubair, A.R.; Alo, O.A. Grey level co-occurrence matrix (GLCM) based second order statistics for image texture analysis. *arXiv* **2024**, arXiv:2403.04038.
10. Chen, X.; Zheng, C.; Yao, H.; Wang, B. Image segmentation using a unified Markov random field model. *IET Image Process.* **2017**, *11*, 860–869. [CrossRef]
11. Almakady, Y.; Mahmoodi, S.; Bennett, M. Adaptive volumetric texture segmentation based on Gaussian Markov random fields features. *Pattern Recognit. Lett.* **2020**, *140*, 101–108. [CrossRef]
12. Hu, S.; Li, J.; Fan, H.; Lan, S.; Pan, Z. Scale and pattern adaptive local binary pattern for texture classification. *Expert Syst. Appl.* **2024**, *240*, 122403. [CrossRef]
13. Farhan, A.H.; Kamil, M.Y. Texture analysis of mammogram using local binary pattern method. *J. Phys. Conf. Ser.* **2020**, *1530*, 012091. [CrossRef]
14. Zhu, L.; Chen, T.; Yin, J.; See, S.; Liu, J. Learning Gabor texture features for fine-grained recognition. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 1–6 October 2023; pp. 1621–1631.
15. Li, Y.; Ge, M.; Zhang, S.; Wang, K. Adaptive segmentation algorithm for subtle defect images on the surface of magnetic ring using 2d-gabor filter bank. *Sensors* **2024**, *24*, 1031. [CrossRef] [PubMed]
16. Lai, X.; Yang, J.; Li, Y.; Wang, M. A building extraction approach based on the fusion of LiDAR point cloud and elevation map texture features. *Remote Sens.* **2019**, *11*, 1636. [CrossRef]
17. Konstantinidis, D.; Stathaki, T.; Argyriou, V.; Grammalidis, N. Building detection using enhanced HOG–LBP features and region refinement processes. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *10*, 888–905. [CrossRef]
18. Ansari, R.A.; Buddhiraju, K.M. Textural classification based on wavelet, curvelet and contourlet features. In Proceedings of the 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, China, 10–15 July 2016; pp. 2753–2756.
19. Ansari, R.A.; Buddhiraju, K.M.; Bhattacharya, A. Textural classification of remotely sensed images using multiresolution techniques. *Geocarto Int.* **2020**, *35*, 1580–1602. [CrossRef]
20. Ansari, R.A.; Buddhiraju, K.M.; Malhotra, R. Urban change detection analysis utilizing multiresolution texture features from polarimetric SAR images. *Remote Sens. Appl. Soc. Environ.* **2020**, *20*, 100418. [CrossRef]
21. Lucas, C.G.; Gilles, J. Demons registration for 2D empirical wavelet transform: Application to texture segmentation. *arXiv* **2024**, arXiv:2409.13075.
22. Su, H.; Chen, J.; Li, Z.; Meng, H.; Wang, X. The fusion feature wavelet pyramid based on FCIS and GLCM for texture classification. *Int. J. Mach. Learn. Cybern.* **2024**, *15*, 1907–1926. [CrossRef]
23. Chen, Z.; Zheng, Y. Research on Extraction Method of Surface Information Based on Multi-Feature Combination Such as Fractal Texture. *J. Geosci. Environ. Prot.* **2023**, *11*, 50–66. [CrossRef]
24. Zhang, K.; Wei, G.; Luo, Y.; Zhao, Y.; Zhao, Y.; Zhang, J. Evaluation of aggregate distribution homogeneity for asphalt pavement based on the fractal characteristic of three-dimensional texture. *Int. J. Pavement Res. Technol.* **2024**, *17*, 577–594. [CrossRef]
25. Tang, Z.; Yan, S.; Xu, C. Adaptive super-resolution image reconstruction based on fractal theory. *Displays* **2023**, *80*, 102544. [CrossRef]
26. Candes, E.J.; Tao, T. Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE Trans. Inf. Theory* **2006**, *52*, 5406–5425. [CrossRef]
27. Donoho, D.L. Compressed sensing. *IEEE Trans. Inf. Theory* **2006**, *52*, 1289–1306. [CrossRef]
28. Yuan, W.; Tian, J.; Hou, B. Image compressed sensing reconstruction algorithm based on attention mechanism. In Proceedings of the International Conference on Computer Vision, Application, and Design (CVAD 2021), Sanya, China, 19–21 November 2021; Volume 12155, pp. 40–45.
29. Xiang, J.; Zang, Y.; Jiang, H.; Wang, L.; Liu, Y. Soft threshold iteration-based anti-noise compressed sensing image reconstruction network. *Signal Image Video Process.* **2023**, *17*, 4523–4531. [CrossRef]
30. Li, Y.; Song, B.; Kang, X.; Du, X.; Guizani, M. Vehicle-type detection based on compressed sensing and deep learning in vehicular networks. *Sensors* **2018**, *18*, 4500. [CrossRef]
31. Emara, H.M.; El-Shafai, W.; Algarni, A.D.; Soliman, N.F.; Abd El-Samie, F.E. A Hybrid Compressive Sensing and Classification Approach for Dynamic Storage Management of Vital Biomedical Signals. *IEEE Access* **2023**, *11*, 108126–108151. [CrossRef]
32. Gao, Z.; Guo, Y.; Zhang, J.; Zeng, T.; Yang, G. Hierarchical perception adversarial learning framework for compressed sensing MRI. *IEEE Trans. Med. Imaging* **2023**, *42*, 1859–1874. [CrossRef]
33. Shakya, A.K.; Vidyarthi, A. Comprehensive study of compression and texture integration for digital imaging and communications in medicine data analysis. *Technologies* **2024**, *12*, 17. [CrossRef]
34. Monika, R.; Dhanalakshmi, S. An optimal adaptive reweighted sampling-based adaptive block compressed sensing for underwater image compression. *Vis. Comput.* **2024**, *40*, 4071–4084. [CrossRef]

35. Altamimi, A.; Ben Youssef, B. Lossless and Near-Lossless Compression Algorithms for Remotely Sensed Hyperspectral Images. *Entropy* **2024**, *26*, 316. [CrossRef]
36. Fu, C.; Du, B.; Huang, X. Hyperspectral image compression based on multiple priors. *J. Frankl. Inst.* **2024**, *361*, 107056. [CrossRef]
37. Fu, C.; Du, B. Remote sensing image compression based on the multiple prior information. *Remote Sens.* **2023**, *15*, 2211. [CrossRef]
38. El-Ashkar, A.M.; Taha TE, S.; El-Fishawy, A.S.; Abd-Elnaby, M.; Abd El-Samie, F.E.; El-Shafai, W. Simultaneous compressed sensing and single-image super resolution for SAR image reconstruction. *Opt. Quantum Electron.* **2023**, *55*, 544. [CrossRef]
39. Xiang, S.; Liang, Q. Remote sensing image compression based on high-frequency and low-frequency components. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5604715. [CrossRef]
40. Gan, H.; Shen, M.; Hua, Y.; Ma, C.; Zhang, T. From patch to pixel: A transformer-based hierarchical framework for compressive image sensing. *IEEE Trans. Comput. Imaging* **2023**, *9*, 133–146. [CrossRef]
41. Xu, J.; Bi, W.; Yan, L.; Du, H.; Qiu, B. An efficient lightweight generative adversarial network for compressed sensing magnetic resonance imaging reconstruction. *IEEE Access* **2023**, *11*, 24604–24614. [CrossRef]
42. Hu, H.; Liu, C.; Liu, S.; Ying, S.; Wang, C.; Ding, Y. Full-Process Adaptive Encoding and Decoding Framework for Remote Sensing Images Based on Compression Sensing. *Remote Sens.* **2024**, *16*, 1529. [CrossRef]
43. Chai, X.; Song, S.; Gan, Z.; Long, G.; Tian, Y.; He, X. CSENMT: A deep image compressed sensing encryption network via multi-color space and texture feature. *Expert Syst. Appl.* **2024**, *241*, 122562. [CrossRef]
44. Peyré, G. Sparse modeling of textures. *J. Math. Imaging Vis.* **2009**, *34*, 17–31. [CrossRef]
45. Böttger, T.; Ulrich, M. Real-time texture error detection on textured surfaces with compressed sensing. *Pattern Recognit. Image Anal.* **2016**, *26*, 88–94. [CrossRef]
46. Wahid, A.; Shah, J.A.; Khan, A.U.; Ahmed, M.; Razali, H. Multi-layer basis pursuit for compressed sensing MR image reconstruction. *IEEE Access* **2020**, *8*, 186222–186232. [CrossRef]
47. Irawati, I.D.; Hadiyoso, S.; Budiman, G.; Fahmi, A.; Latip, R. A novel texture extraction-based compressive sensing for lung cancer classification. *J. Med. Signals Sens.* **2022**, *12*, 278–284. [CrossRef]
48. Yin, Z.; Shi, W.; Wu, Z.; Zhang, J. Multilevel wavelet-based hierarchical networks for image compressed sensing. *Pattern Recognit.* **2022**, *129*, 108758. [CrossRef]
49. Zhang, Y.; Jiang, J.; Zhang, G. Compression of remotely sensed astronomical image using wavelet-based compressed sensing in deep space exploration. *Remote Sens.* **2021**, *13*, 288. [CrossRef]
50. Irawati, I.D.; Budiman, G.; Saidah, S.; Rahmadiani, S.; Latip, R. Block-based compressive sensing in deep learning using AlexNet for vegetable classification. *PeerJ Comput. Sci.* **2023**, *9*, e1551. [CrossRef] [PubMed]
51. Fu, W.; Lu, T.; Li, S. Context-aware compressed sensing of hyperspectral image. *IEEE Trans. Geosci. Remote Sens.* **2019**, *58*, 268–280. [CrossRef]
52. Monika, R.; Samiappan, D.; Kumar, R. Adaptive block compressed sensing-a technological analysis and survey on challenges, innovation directions and applications. *Multimed. Tools Appl.* **2021**, *80*, 4751–4768. [CrossRef]
53. Brodatz, P. *Textures: A Photographic Album for Artists and Designers*; Dover Publications Inc.: New York, NY, USA, 1999.
54. Ji, S.; Wei, S.; Lu, M. Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery data set. *IEEE Trans. Geosci. Remote Sens.* **2018**, *57*, 574–586. [CrossRef]
55. GFDRR Labs. *Open Cities AI Challenge Dataset, Version 1.0*; Radiant MLHub: Washington, DC, USA, 2024. [CrossRef]
56. Daubechies, I.; Meyer, Y.; Lemerie-Rieusset, P.G.; Techamitchian, P.; Beylkin, G.; Coifman, R.; Wickerhauser, M.V.; Donoho, D. Wavelet transforms and orthonormal wavelet bases. *Differ. Perspect. Wavelets* **1993**, *47*, 1–33.
57. Mallat, S. *A Wavelet Tour of Signal Processing*; The Sparse Way: Orlando, FL, USA, 2009; pp. 1–805.
58. Levina, E. *Statistical Issues in Texture Analysis*; University of California: Berkeley, CA, USA, 2002.
59. Haykin, S.S. *Neural Networks and Learning Machines/Simon Haykin*; McMaster University: Hamilton, ON, USA, 2009.
60. Powers, D.M. Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. *arXiv* **2020**, arXiv:2010.16061.
61. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015; Proceedings of the 18th International Conference 2015, Munich, Germany, 5–9 October 2015*; Proceedings, Part III 18; Springer International Publishing: Cham, Switzerland, 2015; pp. 234–241.
62. Zhang, Z.; Liu, Q.; Wang, Y. Road extraction by deep residual u-net. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 749–753. [CrossRef]
63. Li, H.; Qiu, K.; Chen, L.; Mei, X.; Hong, L.; Tao, C. SCAttNet: Semantic segmentation network with spatial and channel attention mechanism for high-resolution remote sensing images. *IEEE Geosci. Remote Sens. Lett.* **2020**, *18*, 905–909. [CrossRef]
64. Wang, L.; Li, R.; Zhang, C.; Fang, S.; Duan, C.; Meng, X.; Atkinson, P.M. UNetFormer: A UNet-like transformer for efficient semantic segmentation of remote sensing urban scene imagery. *ISPRS J. Photogramm. Remote Sens.* **2022**, *190*, 196–214. [CrossRef]
65. Meng, X.; Yang, Y.; Wang, L.; Wang, T.; Li, R.; Zhang, C. Class-guided swin transformer for semantic segmentation of remote sensing imagery. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 6517505. [CrossRef]

66. Chen, J.; Jiang, Y.; Luo, L.; Gong, W. ASF-Net: Adaptive screening feature network for building footprint extraction from remote-sensing images. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 4706413. [CrossRef]
67. Zhao, Y.; Guo, P.; Sun, Z.; Chen, X.; Gao, H. ResiDualGAN: Resize-residual DualGAN for cross-domain remote sensing images semantic segmentation. *Remote Sens.* **2023**, *15*, 1428. [CrossRef]
68. Ansari, R.A.; Mulrooney, T.J. Self-Attention Multiresolution Analysis-Based Informal Settlement Identification Using Remote Sensing Data. *Remote Sens.* **2024**, *16*, 3334. [CrossRef]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.