CURVELET BASED U-NET FRAMEWORK FOR BUILDING FOOTPRINT IDENTIFICATION

Rizwan Ahmed Ansari[1], Winnie Thomas[2]

[1] Symbiosis Institute of Technology, Symbiosis International University, Pune, India – rizwan.vjti@ieee.org
[2] Dept. of Electrical Engineering, Indian Institute of Technology Bombay, India, - winnie.vjti@gmail.com

Commission II, WG II/8

KEY WORDS: Curvelet Transform, Building Identification, Multiresolution Analysis, Semantic Segmentation, Wavelets.

ABSTRACT:

This paper proposes a multiresolution based U-net composite architecture for segmentation of remotely sensed images for building footprint identification. The features derived from curvelet decompositions at different scales are augmented to capture curvilinear discontinuities of the building footprint. This increases the contextual overview of the network as the same data on multiple scales is available for feature extraction and learning. This work further analyses the effects of different multiresolution methods on wavelets and curvelets for decomposition on segmentation performance. The performance is evaluated in terms of precision, recall, F-score, mean intersection over union, overall accuracy, local and global consistency errors. It is found that the proposed method has better class-discriminating power as compared to existing methods and has an overall classification accuracy of 92.4–95.22%. On comparison with the U-Net model performance, it is observed that the proposed network can identify the building areas with higher accuracy and mean intersection over union, the best performance being with curvelet basis of multiresolution analysis.

1. INTRODUCTION

Building footprint extraction from remotely sensed images is vital for the design and regulation of space in urban areas. It has important implications in urban planning, population estimation and topographic map creation. Rapid population increase in urban areas demands rapid infrastructure development which in turn is dependent on robust urban planning. Therefore, there is a growing demand for rapid and automatic building detection techniques. However, automatic building detection has been a long-term challenge in applied remote sensing due to the complex and heterogeneous appearance of buildings in variegated backgrounds.

Detection of buildings from remotely sensed imagery can pose a challenge to semantic segmentation. Techniques involving semantic segmentation of aerial imagery can be classified into two broad categories: traditional methods and deep learning methods. Traditional methods of building detection tend to focus on extracting features that could optimally represent a building. These methods use features ranging from color, texture, shadow, shape, and spatial position relationships of an entity to extract features and then apply either clustering or classification to identify built-up areas (Ansari et al. 2020, Sharma and Singhai 2021, Wang et al. 2021).

Deep neural networks have been used in remote sensing for classification (Liu et al. 2017) and urban analysis (Helber et al. 2019). Fully convolutional networks have shown improved performance for classification (Mullissa et al. 2018, Maggiori et al. 2017). One such network was able to detect different classes and identify their shapes, such as built-up areas, road curvature, and vegetation boundaries. However, it was not capable of detecting small objects and classes with many internal boundaries, as boundaries of these objects may be blurred or improperly oriented, meaning the results are comparatively degraded (Audebert et al. 2016). Recent advances in neural network and deep learning frameworks and Convolutional Neural Networks (CNNs) in particular, have extended its application for building identification, owing to their powerful, nonlinear feature extraction capabilities (Xia et al. 2021, C Li et al. 2021).

Various methods utilizing wavelet-based features in neural networks have also been explored to capitalize on multiscale features of wavelets in the computer vision domain (Liu et al. 2018, Huang et al. 2017). Multiscale convolutional neural networks have also been used for classification (De Silva et al. 2018). Neural networks utilizing multiscale directional features for image semantic segmentation, particularly in the context of remotely sensed image analysis, have only been explored in a limited sense. This work aims to investigate the utility of curvilinear features of curvelet based multiresolution analysis (MRA) in deep learning to identify building footprints from remotely sensed images. The major contribution of this work is to propose a new model based on a set of multiscale curvelet masks as feature maps to include directional information in a deep learning framework with the help of approximation learning. Experimental results show that the proposed curvelet-assisted architecture is more effective than the wavelet-assisted and plain networks in identifying building footprints.

2. MULTIRESOLUTION ANALYSIS

2.1 Wavelet based MRA

The main idea of this work is in the utility of multiresolution analysis in U-net framework. Wavelet based MRA decomposes an image into approximation and detailed subbands by projecting the data onto different basis functions. The following dilation equations with basis functions $\phi(.)$ and $\psi(.)$ are used for decomposition,

Publisher's note: Copernicus Publications has not received any payments from Russian or Belarusian institutions for this paper.

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLVIII-2-W3-2023-15-2023

$$\phi(t) = \sum_n l(n)\phi(2^m t - n) \tag{1}$$

$$\psi(t) = \sum_n h(n)\phi(2^m t - n) \tag{2}$$

where *l[.]* are approximation coefficients and *h[.]* are detailed coefficients of a filter bank, *m* and *n* are scaling dilation and translation indices respectively.

The dilation equations in wavelet transform for MRA use a particular set of basis functions, which are defined by roughly the isotropic functions present at all scales and locations, making it suitable for isotropic features or slightly anisotropic features (Welland, 2003).

A range of other basis functions have been used to extend traditional wavelets, which capture non-linear discontinuities at different scales and aspect ratios to better represent a boundary. A conceptual extension of wavelet-based MRA is a curvelet transform (Welland, 2003), which aims to overcome the representational constraints of wavelets. The curvelet transform is a curvilinear extension to the wavelet transforms in two dimensions constructed using non-separable and directional filter banks. With this set of basis images, it can effectively capture the smooth curves that are the dominant features in remotely sensed images with fewer coefficients. Ansari and Buddhiraju (2019) used curvelet-based texture features for slum identification in remotely sensed images, and it is found that the curvelet-based segmentation provides improved performance over the wavelet-based method.

2.2 Curvelet based MRA

The wavelet transform is optimal at representing straight-line discontinuities in horizontal, vertical and diagonal directions which are rarely observed in remotely sensed images. In order to analyse local line or curve discontinuities, a conventional way is to consider a partition for the image, and then apply the transform in a piece-wise approximation, to the obtained sub-images. Curvelets partition the frequency plane into dyadic (2^j) scales and sub-partition into angular wedges with parabolic aspect ratio. The curvelet transform refines the scale-space viewpoint by adding an extra factor of orientation, and operates by measuring information about an object at specified scales and locations but only along specified orientations.

Curvelet transform works in two dimensions with spatial variable *x*, frequency domain variable *ω*, and the frequency-domain polar coordinates *r* and *θ*. Curvelet transform is defined by a pair of windows, radial window *{W(r)}*, and angular window *{V(t)}*. A polar “wedge” represented by *Uj* is supported by the radial window *{W(r)}* and angular window *{V(r)}.*

2.2.1 Window functions: For constructing the curvelet functions, special window functions are defined which satisfy admissibility conditions. An explicit example is considered here which is representative for all possible choices of window functions being the fundamental to the curvelet construction. For this purpose, the scaled Meyer windows are used (Daubechies, 1992).

$$V(t) = \begin{cases} 1 & ; \quad |t| \le 1/3 \\ \cos\left[\dfrac{\pi}{2}\upsilon(3|t|-1)\right] & ; \quad 1/3 \le |t| \le 2/3 \\ 0 & ; \quad else \end{cases} \tag{3}$$

$$W(r) = \begin{cases} \cos\left[\dfrac{\pi}{2}\nu(5-6\mathrm{r})\right] & ; \quad 2/3 \le r \le 5/6 \\ 1 & ; \quad 5/6 \le r \le 4/3 \\ \cos\left[\dfrac{\pi}{2}\nu(3r-4)\right] & ; \quad 4/3 \le r \le 5/3 \\ 0 & ; \quad else \end{cases} \tag{4}$$

where v is smoothing function defined as

$$\nu(x) = \begin{cases} 0 & ; x \le 0 \\ 1 & ; x \ge 1 \end{cases} \tag{5}$$

In order to obtain smoother functions *W* and *V*, polynomials *$v(x) = 3x^2 - 2x^3$* or *$v(x) = 5x^3 - 5x^4 + x^5$* in *[0, 1]* can be used (Candès and Donoho, 2000). The curvelet elements are obtained as the inverse Fourier transform of a suitable product of the above windows. Therefore, the smoothness of *V* and *W* will ensure a faster decay of the curvelet elements in time domain.

2.2.2 Scaling laws: In the curvelet pyramid, the scale is roughly equal to its length. The anisotropy is increasing with decreasing scales according to a quadratic power law.

This principle gives two additional scaling relations.
- Number of directions is about proportional to the inverse of the scale, and
- Number of micro-locations is about proportional to the inverse of the scale.

Another way of looking at scaling relations within the curvelet transform is to examine the transition from one scale to the next finer scale, i.e. from 2^{-s} to 2^{-s-1}. Each refinement of scale
- *doubles the spatial resolution*; that is, the size of the dyadic squares in the pyramid is reduced by a factor of two (much like wavelet pyramids).
- *doubles the angular resolution*; that is, the number of directions of the anisotropic analyzing elements is increased by a factor of two.

The data are first transformed into the frequency domain by forward discrete Fourier transform (DFT). The transformed data are then multiplied with a set of window functions. The shape of these windows is defined according to the parabolic scaling rule. The curvelet coefficients are obtained by inverse DFT from windowing data. Since the window functions are zero except on support regions of elongated wedges, the regions that need to be transformed by the inverse DFT are much smaller than the original data. On the wrapping curvelet transform, the DFT coefficients on these regions are ‘wrapped’ or folded into rectangular shape before being applied to inverse DFT algorithm. The size of the rectangle is usually not an integer fraction of the size of the original data. This process is equivalent to filtering and subsampling the curvelet subband by rational numbers in two dimensions.

Discrete curvelet transform in the spectral domain utilizes the advantages of fast Fourier transform (FFT). During FFT, both image and curvelet at a given scale and orientation are transformed into the Fourier domain. The convolution of the curvelet with the image in the spatial domain then becomes their product in the Fourier domain. After this step, a set of curvelet coefficients are obtained by applying inverse FFT to the spectral product. This set contains curvelet coefficients in ascending order of the scales and orientations.

The curvelet transform is implemented using a set of directional filters, which are designed using basis functions that have a choice of aspect ratios and directional orientations at multiple scales. The directional filter coefficients effectively capture the anisotropic relationship for curvilinear and disoriented edges. The implementation of a curvelet transform facilitates any level of decomposition, a seamless transition from one scale to another, and faithful reconstruction.

The low-pass filter outputs approximation level information, whereas the band-pass filter extracts the detailed information from a band. The process of decomposition can further be iterated in the low-pass filtered band to extract details of an approximation. These decomposed subbands are augmented with the layers of the U-net to provide multiscale learning, along with directional information.

3. MRA BASED NEURAL ARCHITECTURE

The U-Net is fully convolutional network architecture, containing no fully connected layers which are present in most artificial neural networks. It consists of two units- the encoder and the decoder. The encoder is the down-sampling path made up of convolutional and max-pooling layers, which results in the activation maps to get compressed successively while capturing contextual information at every level of compression. The decoder comprises up-sampling layers and convolutional layers which expand the activation maps to get back the dimensions of the original image. The main intuition behind this network is that the down-sampling path increases the receptive field of the network, and the information lost in this process is given to the up-sampling path for the reconstruction of the original structure of the image. This is achieved by the skip connections which direct the higher-level features from the encoder directly to the decoder layers. Hence the learned features during down-sampling are utilized in the up-sampling process which results in smoother boundaries than an ordinary fully connected network would provide. Figure 1 describes the proposed curvelet based U-net architecture.

In this paper, a version of the U-net (Ronneberger et al. 2015) is used by augmenting curvelet subbands of different size at different scales. During feature extraction, there are four steps, with the last three including several subbands. The feature maps in the same levels have the same size, while the feature maps in the following level are half that of the previous level. For a 3-level decomposition, there are 16, 8, and 4 subbands respectively. The expansive part aims to extract feature maps for informal settlements using contourlet masks. The number of stages in contracting and expansive parts is the same. Having a convolutional layer followed by a max-pooling layer helps in gathering contextual information present at each level of decomposition in terms of generating activation functions. The decoder expands these activation functions with the help of the up-sampler and convolutional units to obtain the original size of a band. The central aim is to enhance the receptive field of the model using the down-sampler. The residual information in the process is fed to the up-sampler for faithful reconstruction. This is attained by the skip connections in the network, while the features learned during down-sampling are used in the up-sampling part. In turn, this mechanism provides smoother edges than other fully connected convolutional networks.

Building footprint identification is considered as a binary classification. For training, logistic regression is used by optimizing the energy function. A gradient decent algorithm is used to minimize the error function. Both the softmax and cross-entropy functions are considered for the error function. The softmax layer outputs two lines as a probability indicator for informal settlements and rest of the classes. The last layer is a convolutional layer measuring 1x1, which is used to transform the features into two classes for the pixel under consideration. The concatenation in the expansive segment is able to learn the features at multiple scales. The feature learning at multiple scales enhances the ability to capture different properties of the classes and improves the classification accuracy.

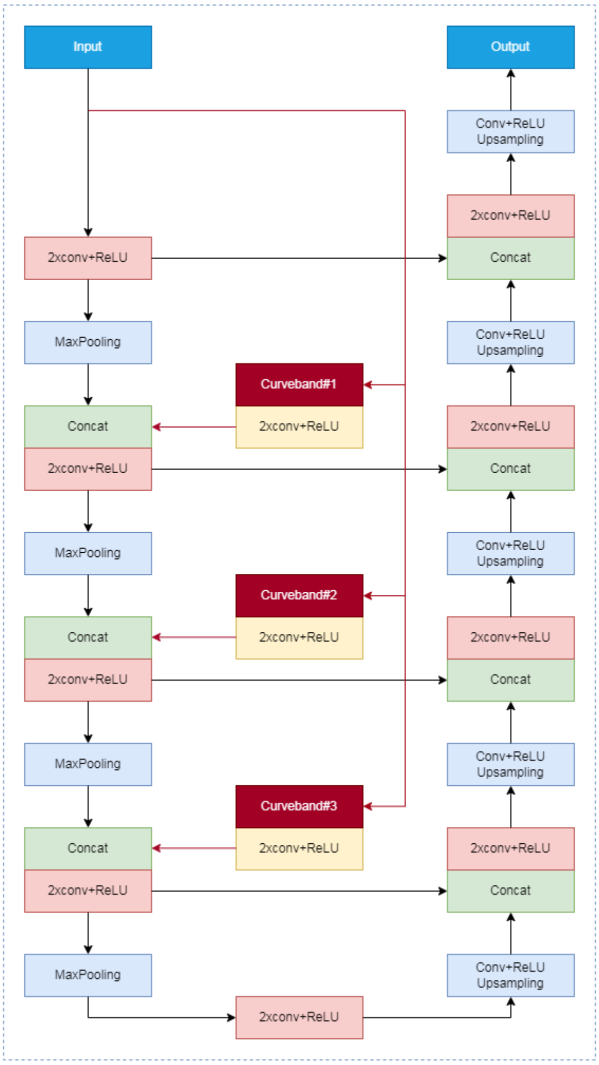


Figure 1. Proposed curvelet based U-net architecture (Conv: Convolution layer; ReLU: Rectified Linear Unit; MaxPooling: Maximum Pooling; Concat: Concatenation; Curveband#x: curvelet subband at decomposition level x)

4. RESULTS AND ANALYSIS

To demonstrate the robustness of the proposed algorithm, we evaluated the performance on the OpenCities dataset (GFDDR labs, 2020) which consists of thousands of building footprints from various cities and regions in Africa extracted from OpenStreetMap. The spatial resolution of the aerial imagery varies from region to region, varying from 0.03m x 0.03m - 0.2m x 0.2m. Out of ten cities in the dataset, we used aerial imagery from three cities (Accra, Dares Salaam, Zanzibar) so as to have a dataset consisting of densely and sparsely populated regions alike with varying resolutions. This enables us to demonstrate that the method is not biased to a specific resolution and/or a particular building structure or clutter size since the imagery in the dataset is diverse, ranging from countryside to industrial areas consisting of multifarious architecture styles.

To evaluate the semantic segmentation performance of the proposed model, we used mean intersection over union (mIoU), precision, recall (R), F1-score (F1), and overall pixel accuracy (OA) as performance metrics. We compare and contrast our method with several other state-of-the-art architectures including U-net, random forest classifier, support vector machines and fully connected network (FCN-32s). Table 1 details the quantitative semantic segmentation results averaged across multiple images of varying spatial resolutions and building clutter taken over a large study area of all the networks implemented for comparative analysis. Figure 2 shows qualitatively the comparative performance of the proposed methodology contrasted against the state-of-the-art architectures.

Model	mIoU	P	R	F1	OA(%)
U-Net	0.91	0.90	0.89	0.92	92.03
FCN-32s	0.89	0.85	0.86	0.87	91.17
Random Forest	0.84	0.87	0.88	0.89	90.07
SVM	0.86	0.84	0.79	0.89	89.74
Wavelet U-net	0.92	0.91	0.90	0.91	93.24
Proposed U-Net	0.94	0.92	0.92	0.93	95.22

mIoU: mean intersection over union, P: precision, R: recall, F1: F1 score, OA: percentage overall accuracy

Table 1. Performance Comparison

	Image#1		Image#2	
Model	LCE	GCE	LCE	GCE
U-Net	0.41	0.43	0.39	0.42
Wavelet U-net	0.085	0.089	0.086	0.089
Proposed U-Net	0.070	0.078	0.072	0.080

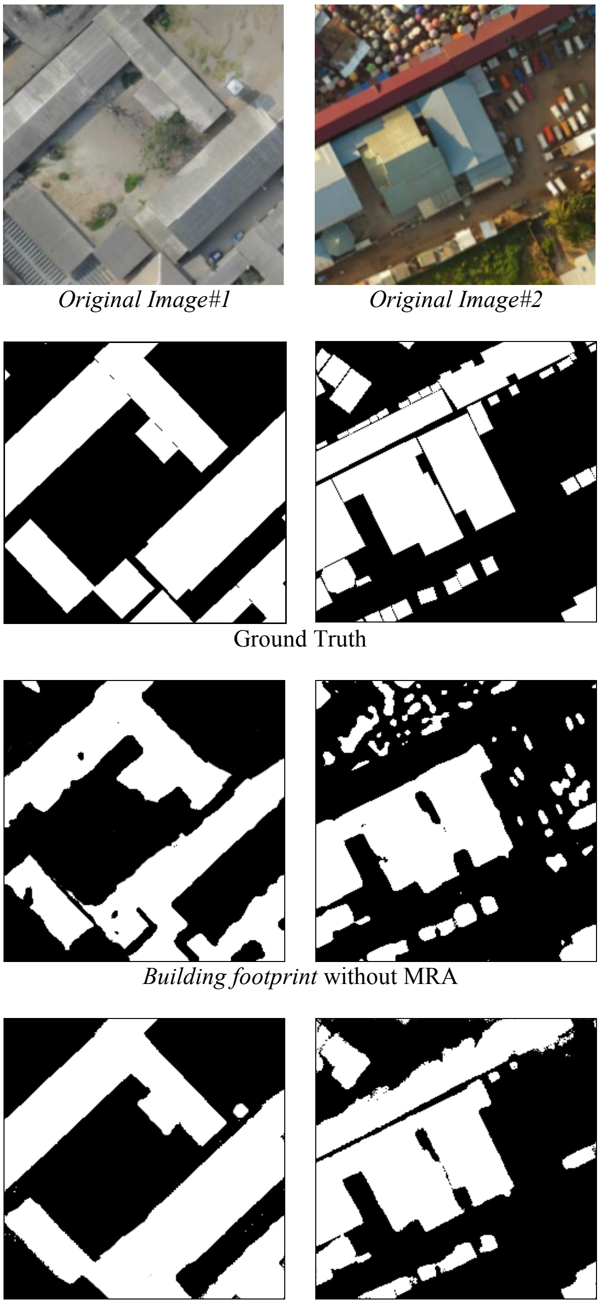
LCE: Local consistency error, GCE: Global consistency error

Table 2. Degree of matching performance

In order to assess the effectiveness of different MRA features apart from visual interpretation, local and global consistency errors (LCEs and GCEs) are computed as quantitative measures to evaluate the degree of matching between segmentation output and the reference site (Table 2).

GCE forces all local refinements to be in the same direction and assumes that one of the segmentations must be a refinement of the other. LCE and GCE quantify the degree of matching between segmentation results obtained from different MRA methods and the reference segmentation window generated from human visual interpretation. The lower values of LCE and GCE demonstrate higher degree of matching for curvelet (0.07, 0.078) and wavelet (0.085, 0.089) features when compared to non-MRA (0.39, 0.42) segmentation.

The experimental results for the curvelet based method exhibited good performance in terms of both visual interpretation and feature discrimination and are sufficiently robust against random pixels while preserving spatial arrangement. An overall classification accuracy of 92.4–95.22% is achieved with proper boundary shapes using curvelet method.



Original Image#1 *Original Image#2*

Ground Truth

Building footprint without MRA

Proposed Method

Figure 2. Building footprint identification

5. CONCLUSIONS

In this work, we developed a curvelet based multiresolution analysis approach in U-net framework for the building footprint extraction. The method progressively combines subbands of curvelet decompositions at various scales to extract different disoriented details, which are key manifestations of the building footprints in remotely sensed images. The proposed algorithm is tested on OpenCities images of covering different regions of formal and informal urban settlements. The results were compared with plain U-net and wavelet-assisted U-net models. The performance was evaluated based on the visual interpretation, precision, recall, F-score, mIoU, and overall accuracy of these methods. The results showed that multiscale curvelet subbands in the proposed U-net yielded better accuracy and boundary continuity than the prior proposals. The improved performance was because of the ability of the curvelet transform to capture curvilinear features of linear and nonlinear discontinuities when compared with plain U-net and wavelet-assisted U-net models. This approach may be extended to train other architectures since there is no network-dependent modification required. Future scope includes evaluating the applicability of the method for other remote sensing tasks like change detection, land use and land cover classification, among others.

ACKNOWLEDGEMENTS

The authors would like to thank the anonymous reviewers for their valuable suggestions.

REFERENCES

Audebert, N.; Saux, B.L.; Lefèvre, S. Semantic Segmentation of Earth Observation Data Using Multimodal and Multi-scale Deep Networks. In Proceedings of the Computer Vision—ACCV 2016, Taipei, Taiwan, 20–24 November 2016; pp. 180–196

Ansari, R.A.; Buddhiraju, K.M. Textural segmentation of remotely sensed images using multiresolution analysis for slum area identification. Eur. J. Remote Sens. 2019, 52 (Suppl. 2), 74–88.

Ansari, R. A., Buddhiraju, K. M., & Malhotra, R. (2020). Urban change detection analysis utilizing multiresolution texture features from polarimetric SAR images. *Remote Sensing Applications: Society and Environment*, *20*, 100418.

C. Li et al., "Attention Residual U-Net for Building Segmentation in Aerial Images," 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, 2021, pp. 4047-4050.

Candès EJ, Donoho DL. 2000. December. Curvelets, multiresolution representation, and scaling laws. In International Symposium on Optical Science and Technology (pp. 1–12). International Society for Optics and Photonics.

Daubechies, Ingrid. *Ten lectures on wavelets*. Society for industrial and applied mathematics, 1992.

De Silva, D.D.N.; Fernando, S.; Piyatilake, I.T.S.; Karunarathne, A.V.S. Wavelet based edge feature enhancement for convolutional neural networks. In Eleventh International Conference on Machine Vision (ICMV 2018); International Society for Optics and Photonics: Munich, Germany, 2019; Volume 11041, p. 110412R.

GFDRR Labs (2020). "Open Cities AI Challenge Dataset", Version 1.0, Radiant MLHub. 20 Dec 2021

J. Xia, N. Yokoya, B. Adriano, L. Zhang, G. Li and Z. Wang, "A Benchmark High-Resolution GaoFen-3 SAR Dataset for Building Semantic Segmentation," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 14, pp. 5950-5963, 2021.

Helber, P.; Bischke, B.; Dengel, A.; Borth, D. Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification. IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens. **2019**, 12, 2217–2226.

Huang, H.; He, R.; Sun, Z.; Tan, T. Wavelet-srnet: A wavelet-based cnn for multi-scale face super resolution. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 1689–1697.

Liu, P.; Zhang, H.; Zhang, K.; Lin, L.; Zuo,W. Multi-level wavelet-CNN for image restoration. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 19–21 June 2018; pp. 773–782

Liu, Q.; Hang, R.; Song, H.; Li, Z. Learning multiscale deep features for high-resolution satellite image scene classification. IEEE Trans. Geosci. Remote Sens. **2017**, 56, 117–126.

Maggiori, E.; Tarabalka, Y.; Charpiat, G.; Alliez, P. Convolutional Neural Networks for Large-Scale Remote-Sensing Image Classification. IEEE Trans. Geosci. Remote Sens. **2017**, 55, 645–657.

Mullissa, A.G.; Persello, C.; Tolpekin, V. Fully Convolutional Networks for Multi-Temporal SAR Image Classification. In Proceedings of the IGARSS 2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 6635–6638.

Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional networks for biomedical image segmentation. In Proceedings of the Medical Image Computing and Computer Assisted Intervention (MICCAI), Munich, Germany, 5–9 October 2015; pp. 234–241

Sharma, Deepa, and Jyoti Singhai. "An unsupervised framework to extract the diverse building from the satellite images using Grab-cut method." Earth Science Informatics 14.2 (2021): 777-795.

Wang, Chao, et al. "Automatic Building Detection from High-Resolution Remote Sensing Images Based on Joint Optimization and Decision Fusion of Morphological Attribute Profiles." Remote Sensing 13.3 (2021): 357.

Welland G, editor. 2003. Beyond wavelets (Vol. 10). New York: Academic Press.