Automatic 3D Multiple Building Change Detection Model Based on Encoder–Decoder Network Using Highly Unbalanced Remote Sensing Datasets

Masoomeh Gomroki, Mahdi Hasanlou, *Member, IEEE*, and Jocelyn Chanussot, *Fellow, IEEE*

***Abstract*—3-D building change detection (CD) methods detect more accurate multiple change maps than 2-D ones. Recent technologies, such as unmanned aerial vehicle (UAV) systems and dense image matching have made it much easier to obtain 3-D data nowadays. Developing a solution which produces an accurate map of multiple building changes, including *unclassified*, *no building change*, *newly built*, *demolished*, and *taller*, at an acceptable speed is a challenging issue. In this article, we address a novel 3-D building CD method based on an encoder–decoder network to detect accurate multiple changes maps automatically, in the presence of highly unbalanced remote sensing datasets. The proposed method consists of three main parts: the preprocessing and mixed augmentation (MA) step; the encoder–decoder network training; and finally the prediction step. The data are augmented by the MA method to manipulate highly unbalanced datasets. The encoder–decoder network is constructed by the Yolov7 network as the encoder path and the decoder path equipped with the convolutional layers of modified Unet). Two datasets are used in this article. The first dataset is the point clouds and orthophotos obtained from the UAV of Mashhad City in 2011 and 2016. The second dataset consists of stereo images of the GeoEye-1 satellite and the point clouds obtained from dense image matching of Tehran city in 2009 and 2013. The results show that the proposed method achieved accuracy and kappa coefficients above 94% and 0.90 for both datasets, respectively.**

***Index Terms*—3-D multiple building change detection (CD), convolutional layers of modified Unet (CLMUnet), fully automatic, unbalanced remote sensing dataset, Yolov7.**

I. Introduction

AUTOMATIC change detection (CD) has recently been recognized as an essential topic in remote sensing and photogrammetry [1]. Buildings are the main elements of urban areas so detecting building changes plays a critical role in providing comprehensive spatial information on rural-urban transition, illegal constructions, city developments [2], [3], [4], [5] and disaster management [5], [6]. Previous researches presented CD methods based on only spectral information of remote sensing images have weaknesses such as relief displacement, shadow presence, occlusion and spectral variation of buildings [5]. Make use of UAVs, laser scanners, dense image (stereo) matching, digital terrain models (DTMs) and digital surface models (DSMs) building height information was created so that 3-D CD in buildings began [5].

As discussed in [7], traditional methods of 3-D CD are classified into two broad categories: geometric comparison and geometric-spectral analysis. There are three types of geometric comparison methods: height differencing; Euclidean distance height differencing; and projection-based differencing [7]. The change map in the height differencing method is obtained from the difference between two DSMs [8], [9], [10], [11], [12]. The height differencing method is simple to implement and effective for large-scale CDs, but it is sensitive to misregistration and image matching errors and it can be used for 2.5-D levels [5], [7]. The Euclidean distance of two 3-D surfaces is calculated in Euclidean distance height differencing [13], [14]. Its stability against small errors in registration for top-view 3-D data as well as its application to full 3-D data comparison is two advantages of this method, but its correspondence search is time consuming and its implementation is complicated [7]. The projection-based differencing method measures geometric differences. In this method, the correlation of stereo images from one epoch is calculated using point cloud or DSMs of other epochs and then the correlations of these two data are compared and the amount of spectral inconsistency is calculated [15], [16]. In homogeneous areas this method may have missing detections and the accuracy of 3-D data is important [5], [7].

Geometric-spectral analysis methods are postrefinement; direct feature fusion; and postclassification [7]. The postrefinement method improves geometric comparison results, such as height differencing by utilizing geometric and spectral features [17], [18], [19], [20]. Although this method is flexible and efficient, the results depends on the geometric comparison and the missing changes cannot be recovered in the subsequent steps [5], [7]. To calculate simultaneously the changes in spectral and geometric features, the direct feature fusion method is considered [21], [22]. The combination of geometric and radiometric information, as well as the simultaneous using of different information bands without the necessity of the algorithm

Manuscript received 30 July 2023; revised 15 September 2023; accepted 24 October 2023. Date of publication 30 October 2023; date of current version 23 November 2023. *(Corresponding author: Mahdi Hasanlou.)*

Masoomeh Gomroki and Mahdi Hasanlou are with the School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran 1439957131, Iran (e-mail: masoomeh.gomroki@ut.ac.ir; hasanlou@ut.ac.ir).

Jocelyn Chanussot is with CNRS, Grenoble INP, GIPSA-Lab, University Grenoble Alpes, 38000 Grenoble, France, and also with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China (e-mail: jocelyn.chanussot@grenoble-inp.fr).

Digital Object Identifier 10.1109/JSTARS.2023.3328561

improvement are advantages of this method, but determining the appropriate parameters for the fusion is a major challenge [5], [7]. The postclassification method classifies objects on each dataset separately and then compares the results between labels [23], [24], [25]. This method improves classification and object detection accuracy compared with other traditional methods, but the accuracy of the final results depends on the accuracy of the classification [5], [7].

In addition to traditional methods, deep learning methods also introduced and applied for 3-D CDs, as they have previously been used in the 2-D CDs (e.g., applications in urban land use and urban land cover [26], [27], [28], [29], [30], [31], [32], building CD [33], [34], [35], [36], disaster monitoring [37], [38], urban planning [39], [40] and resource survey [41], [42], [43]), and obviously deep learning studies in 3-D CDs is in its early stages. Zhang et al. [44] used Lidar data of point clouds obtained from image matching for the 3-D binary CD in a deep learning network. Pang et al. [5] created a 3-D change map using a deep convolution network in addition to a graph-based and simultaneous segmentation method. Yew and Lee [45] applied a convolution network to balanced point clouds obtained from "structure from motion" of two times for urban CD. In another study, Yadav et al. [46] used Lidar point clouds for 3-D building CD in the Unet network. Lian et al. [47] detected changes in buildings using two-time DSMs in an end-to-end convolution network with five convolution layers. Nagy et al. [48] used CNN network architecture to detect changes in urban areas. Blocks of a Siamese network, Unet network and a transformer were used in that research to provide a map of binary changes only in urban streets. de Gélis et al. [49] used a Siamese network composed of kernel point convolution blocks for 3-D binary CD on urban point clouds. Mohammadi and Samadzadegan [50] used CNN only to extract 3-D features from stereo images of point clouds and then presented the 2-D and 3-D changes of buildings in urban areas. Amini Amirkolaee and Arefi [51] used a convolutional network only to estimate DSM and detected 3-D changes in urban areas. It is noticeable that in all of these researches balanced data has been assumed.

Deep learning methods in 3-D multiple CD are a new topic and have not reached the same depth as research on 2-D CD. Considering the necessity and importance of 3-D building CD in this article, we proposed an efficient and accurate method with an acceptable speed that can provide a map of multiple building changes. Two highly unbalanced datasets are considered to detect 3-D building changes. The first one is the point clouds and orthophotos obtained from the UAV of Mashhad City in 2011 and 2016. The second dataset consists of stereo images of the GeoEye-1 satellite and the point clouds obtained from dense image matching of Tehran city in 2009 and 2013.

Highly unbalanced data has always been a significant and difficult issue for most researches, particularly in deep learning and machine learning. In our method this challenge has been resolved by mixed augmentation (MA) and unbalanced data has been properly distributed into the training network. Then, an encoder–decoder network is employed. To extract features in the encoder path we use the Yolov7 network [52] which has a high level of accuracy and speed compared to other similar networks. Semi-transfer learning technique is utilized in the sense that the Yolov7 was pretrained by the MS COCO dataset [27]. On the other hand, the convolutional layers of modified Unet (CLMUnet) are used for the decoder path. By modified Unet we mean that instead of a fixed-size kernel, a variable-size kernel is used, making this network more capable for multiple CDs.

The main contributions of this article are as follows.

1) Introduce an encoder–decoder network in 3-D multiple CD by utilizing the combination of Yolov7 and CLMUnet.
2) Get the advantages of a transfer learning technique in the encoder path (Yolov7 pre-trained by the MS COCO dataset) to reduce the training time and the limitations of GPU capacities.
3) Manipulate the highly unbalanced datasets in 3-D CD by MA technique to increase the overall accuracy and kappa coefficient.
4) Propose a deep learning network to completely automated the 3-D multiple CD of buildings in urban areas.

The rest of this article is organized as follows. The studied datasets are discussed in Section II. The proposed method is explained in Section III. The experimental results obtained from this method and the comparisons are expressed in Section IV. The discussion of the results is presented in Section V. Finally, Section VI concludes this article.

II. Material and Datasets

A. UAV Images and Point Clouds Dataset (the Densely Built Mashhad City)

The first dataset used in this article is the images and the point clouds taken by UAV in two epochs 2011 and 2016 from a densely built urban area in the city of Mashhad, Iran. The geometric structure and different heights of the buildings, as well as the different spectral structures of their roofs express the diversity and complexity of this area. This dataset includes the DSMs obtained from the point clouds and the orthophotos at both times, depicted in Fig. 1. The spatial resolution of this data (10 cm) enables us accurate identification and CD in buildings. To generate ground truth the orthoimages and their corresponding Google Earth images are used. The polygons of buildings are drawn in the global mapper software and the different layers of "*no building change*, *newly built*, *demolished*, and *taller*" are separated and different numbers are assigned to them so that the ground truth image is obtained as shown in Fig. 1(e). All these image datasets can be downloaded from rslab.ut.ac.ir.

B. GeoEye-1 Satellite Stereo Image Dataset (Tehran City Development)

The second dataset examined in this article is from the 22nd district of the city of Tehran. This is one of Tehran's newest and largest geographical urban areas located northwest

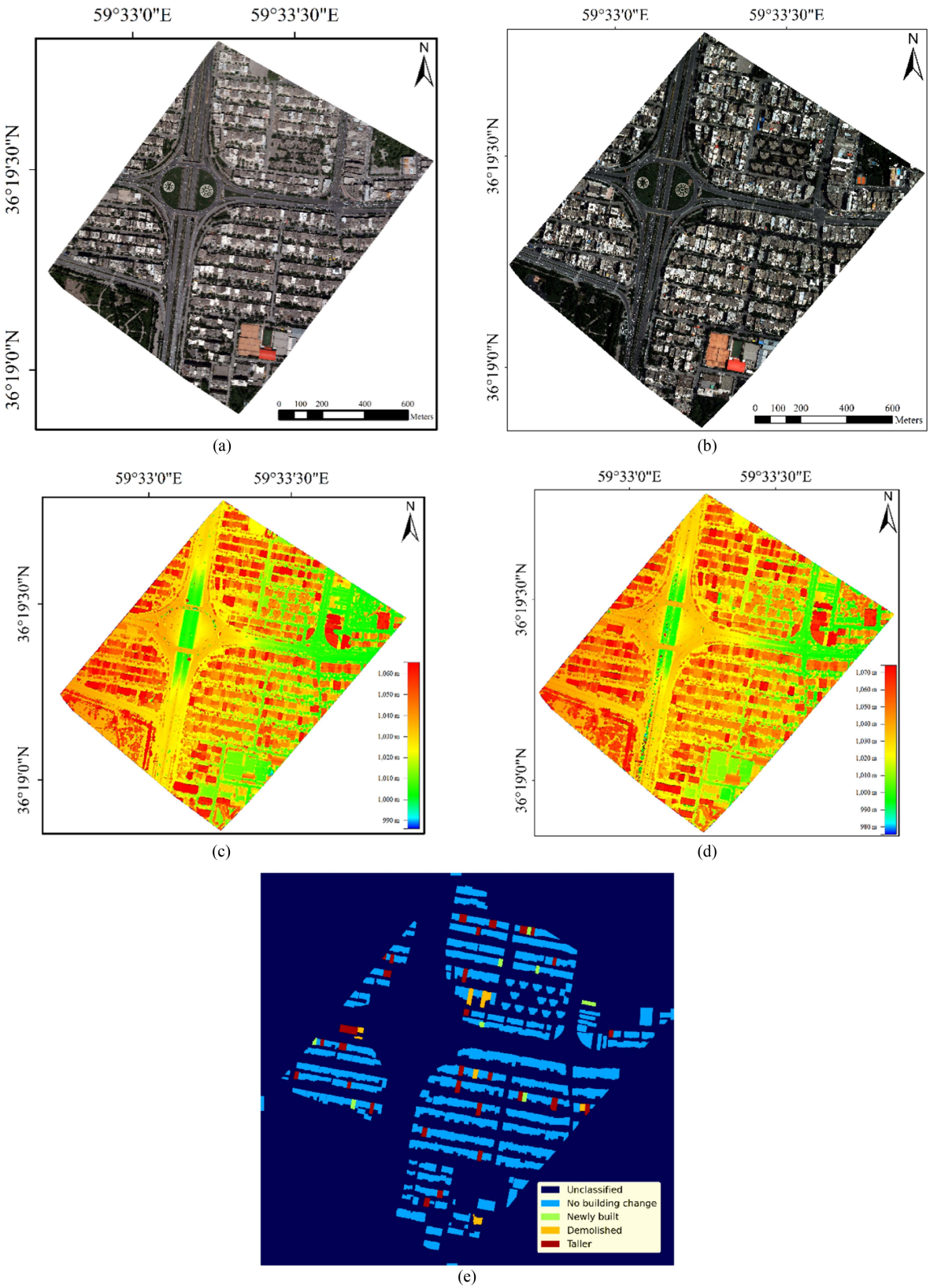


Fig. 1. UAV of the Mashhad city dataset includes: (a) RGB time1; (b) RGB time2; (c) DSM time1; (d) DSM time2; and (e) ground truth.

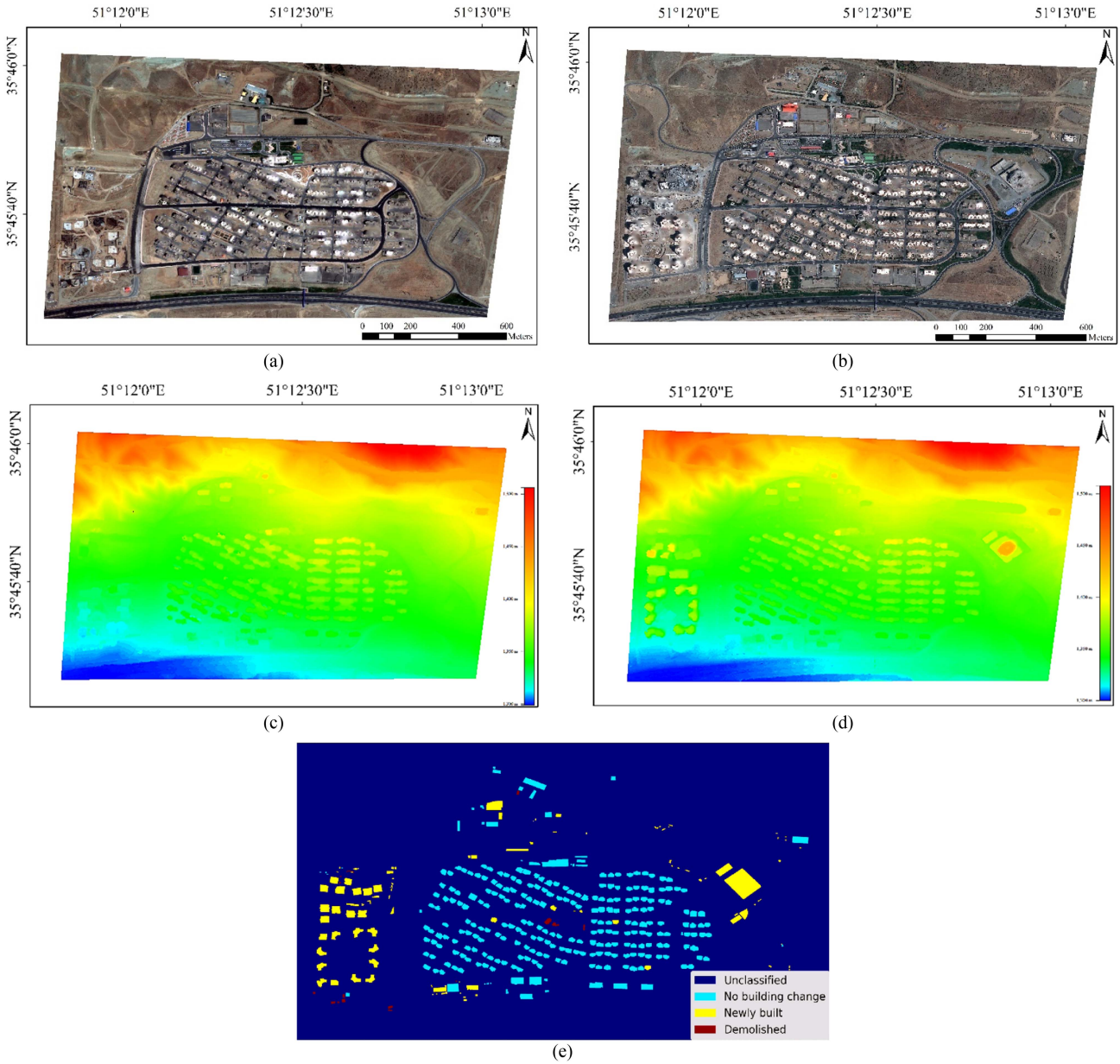


Fig. 2. GeoEye-1 satellite stereo images of Tehran city include: (a) RGB time1; (b) RGB time2; (c) DSM time1; (d) DSM time2; and (e) ground truth.

of it which has undergone significant changes in recent years due to urban development such as the construction of residential towers, commercial centers, shopping malls and local access roads. This dataset includes two stereo-images of the GeoEye-1 satellite with a spatial resolution of 0.5 m for epochs 2009 and 2013. The semiglobal matching technique [53] is employed to generate the point clouds from these stereo-images.

Fig. 2 exhibits the satellite images of both times and their DSMs generated from the point clouds. The ground truth is generated by using the satellite images of both times and their Google Earth images. The polygons of buildings are drawn and labeled in the Global Mapper in the same way as the first dataset. The ground truth is depicted in Fig. 2(e). Table I gives the two datasets.

III. Proposed Method

The method for 3-D building CD established in this article is investigated thoroughly in this section. This method basically consists of three steps: the preprocessing and MA; the encoder–decoder network training for 3-D building CD; and the prediction step using the trained network.

Initially, to overcome the highly unbalanced problem, the data is augmented by the MA method and then the preprocessing step is performed to prepare the data as the input of the encoder–decoder network, as well as to partition it into the training, the validation and the test parts. The training and the validation parts are then considered as the input of the encoder–decoder network in the second step. This network includes Yolov7 pretrained by

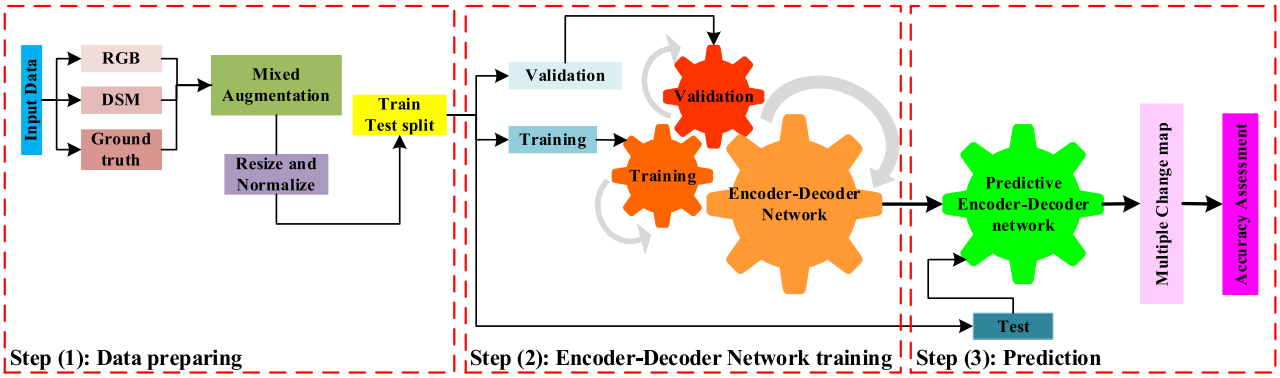


Fig. 3. Flowchart of the established method.

TABLE I
INFORMATION OF DATASETS USED IN THIS ARTICLE

Datasets	Time	Input networks	Spatial Resolution (m)	Study area (km^2)
UAV of Mashhad city	2011	RGB time1 DSM time1	0.1	47.1
	2016	RGB time2 DSM time2		
GeoEye-1 satellite stereo images of Tehran city	2009	RGB time1 DSM time1	0.5	2.5
	2013	RGB time2 DSM time2		

the MS COCO dataset as the encoder path and CLMUnet as the decoder path. Finally, the trained network will be evaluated. Fig. 3 depicts the general flowchart of this method.

A. Preprocessing and Mixed Augmentation

The classification and segmentation of unbalanced data have always been regarded as a challenging issue [54]. The weight of each class is calculated by [55]:

$$W_i = \frac{n_{\text{samples}}}{n_{\text{classes}} \times F_i}. \tag{1}$$

In this equation W_i is the weight of each class, n_{samples} is the amount of the whole data, n_{classes} is the number of classes and F_i is the frequency of each class in the dataset. As the ground truth in Fig. 1(e) shows, a highly unbalanced dataset is dealt with in this article. For example, in the Mashhad UAV dataset the weights of *Unclassified*, *no building change*, *newly built*, *demolished*, and *taller* classes are 0.32, 0.59, 30.50, 42.43 and 8.56, respectively, so *newly built* and *demolished* classes are rare. Similarly, the second dataset also is highly unbalanced according to the *demolished* class, since the weights of *unclassified*, *no building change*, *newly built* and *demolished* classes are 0.28, 3.17, 8.01, and 170.61, respectively. In previous researches different loss functions or a combination of them have been used to tackle the problem of data unbalancing [56], but those methods are not applicable in our case. An effective solution is to use the MA method emphasized on the rare classes which unbalance the data. In [57], MA methods were classified into: linearity-based and nonlinear. One of the nonlinear methods is the horizontal concat method which is represented by

$$Y = \lambda \times X_1 + (1 - \lambda) \times X_2. \tag{2}$$

In this method λ fraction of the image $X_1$is concatenated to $(1 - \lambda)$ fraction of the image X_2 to produce the augmented image Y. We generate three MA images by considering λ equals to 0.3, 0.5 and 0.7. These MA stages are applied to the images which contain *demolished* and *newly built* classes in the first dataset and *demolished* class in the second one. Fig. 4 shows one MA data result and their corresponding ground truths in the first dataset.

In addition to the MA method, all images in every dataset are also augmented by $+90°$ and $+180°$ rotations. All images are normalized and resized and then they partition into the training, the validation and the test parts.

B. Proposed Encoder–Decoder Network Architecture for 3-D CD

The encoder–decoder network of this article consists of two networks: Yolov7 and CLMUnet. Yolov7 was pre-trained by the MS COCO dataset and since it is used only in the encoder part, the technique is called "the semitransfer learning method" [27]. The convolutional layers of conventional Unet are modified in the decoder part to handle the complexity of our problem and the studied data. The resulting network has advantages over previous 3-D CD methods. Despite of the former researches in which the deep learning networks were employed to perform only a few steps of the 3-D CD process, our constructed encoder–decoder network performs the entire 3-D CD process automatically. Furthermore, this method can efficiently and accurately extract the features of the input data at an appropriate speed. Our modification to the structure of the convolutional blocks in Unet and using kernels with different sizes handle the diversity and complexity of urban buildings so that final multiple change maps are generated more accurately. Finally, the network employs the focal loss function which has proper performance on unbalanced data classification [58].

1) Encoder Path: In the case of 3-D multiple building CDs in the presence of a highly unbalanced dataset, we need

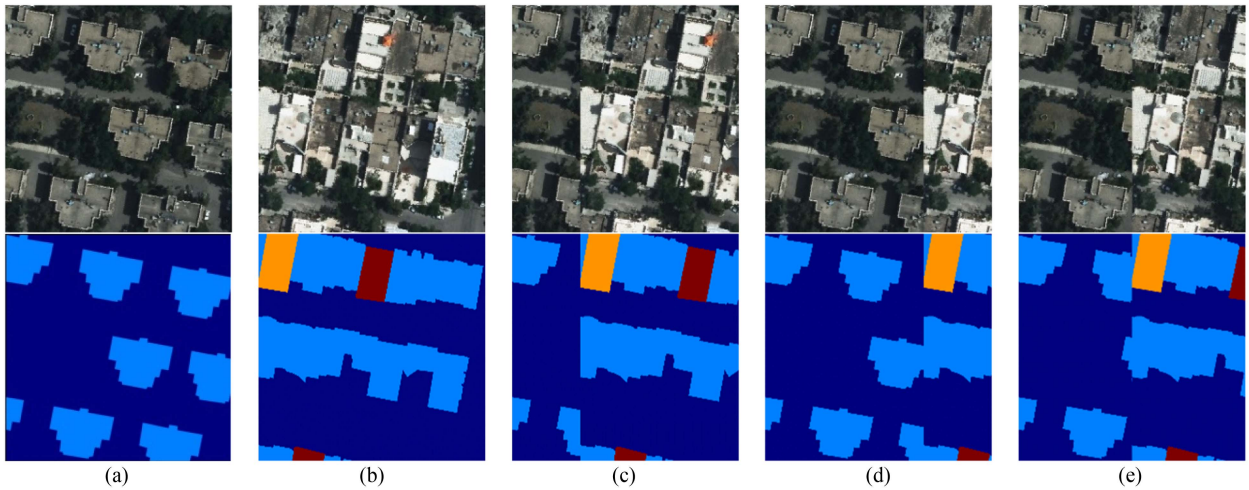


Fig. 4. MA and their corresponding ground truth: (a) first image; (b) second image; (c) MA image produced by 0.3 of the first image and 0.7 of the second image; (d) MA image produced by 0.7 of the first image and 0.3 of the second image; and (e) MA image produced by 0.5 of the first image and 0.5 of the second image.

to construct a single-stage object detector encoder network that not only has good performance in acceptable speed but also it can effectively and accurately extract the features of the images and could be pre-trained. To achieve this goal, we look for and examine more than 15 networks including transformer-based networks and finally, blocks of Yolov7 assist us to achieve our encoder requirements. We developed the encoder path of the proposed network by utilizing some blocks of Yolov7 to profit from its advantages and avoid its weaknesses.

Redmon et al. [59] were the first to use the Yolo network for object detection. Yolo models are commonly used to detect and classify objects by only looking at an image or video once [60]. Yolo networks consist of three main parts: backbone; neck; and head [60]. The backbone extracts the features from the input image, the neck generates feature pyramids and the head (or the predictor) displays the final results of the network [60]. The Yolov7 network was introduced in 2022 and it showed the best speed and accuracy compared to other known networks for object detection [52]. The Yolov7 network modules include CBS, ELAN, SSPPCSPS, and MP. CBS includes a convolutional layer, batch normalization and the Silu activation function [52]. The Yolov7 network was the first to introduce ELAN architecture which enables the network to use expand, shuffle and merge cardinality techniques to effectively train the model while preserving the original gradient route [60]. Down-sampling is accomplished by the MP structure [61]. Fig. 5 depicts the Yolov7 network structure in detail.

2) Decoder Path: Unet is a U-shaped network that was first used to segment medical images by Ronneberger [62]. The Unet network has two paths: encoder and decoder. The encoder path extracts deep features from the input, while the decoder path uses transpose convolution to determine the exact position of the features [63]. Each convolutional layer in the Unet network takes the form of Fig. 6(a) which includes a convolution with kernel size 3×3 and Relu activation function alternately. In each layer of the MUnet network, a batch normalization layer has been inserted in addition to the convolutions with kernel sizes of 3×3 and 5×5 [see Fig. 6(b)]. In this article, the convolutional layers of the MUnet network are used as the decoder path. Gathering all the above, Fig. 7 depicts the network architecture used in this article.

C. State-of-the-Art Methods

Taking the advantages of encoder–decoder deep learning networks is one of the contributions of this article, so networks such as Yolo families, EfficientNet families and some transformer-based models were investigated. Transformer models have received a lot of attention in recent deep-learning researches. We implemented and compared the performance of some state-of-the-art networks that have a pyramid and hierarchical design in which their blocks can be partitioned. Because of the GPU limitations, all of the networks should be pretrained. Furthermore, only versions which are executable based on the system limitations are considered. These networks will be introduced briefly in the following.

1) *EfficientNet V2 [64]:* This network consists of a set of training-aware neural architectures that are intended to improve training speed and parameters. Versions B0, B1, B3, and T are supposed here because they have fewer parameters than the others. This network was pretrained by the ImageNet dataset.
2) *Efficientformer V2 [65]:* This network is based on vision transformer models, but it is faster and more efficient. The S0, S1, S2, and L versions of this model which had fewer parameters are assumed.
3) *TinyViT [66]:* TinyViT networks resolve the problem of having a large number of parameters in vision transformer-based models. The main idea is based on transformer knowledge from a large pre-trained model to a small one. These networks were pre-trained by ImageNet.
4) *YoloX [67]:* Yolo networks are used for object detection, as detailed in Section III-B.1. The Yolox network is one of the fastest and most effective networks in the Yolo family. This network was pre-trained by the MS COCO dataset.

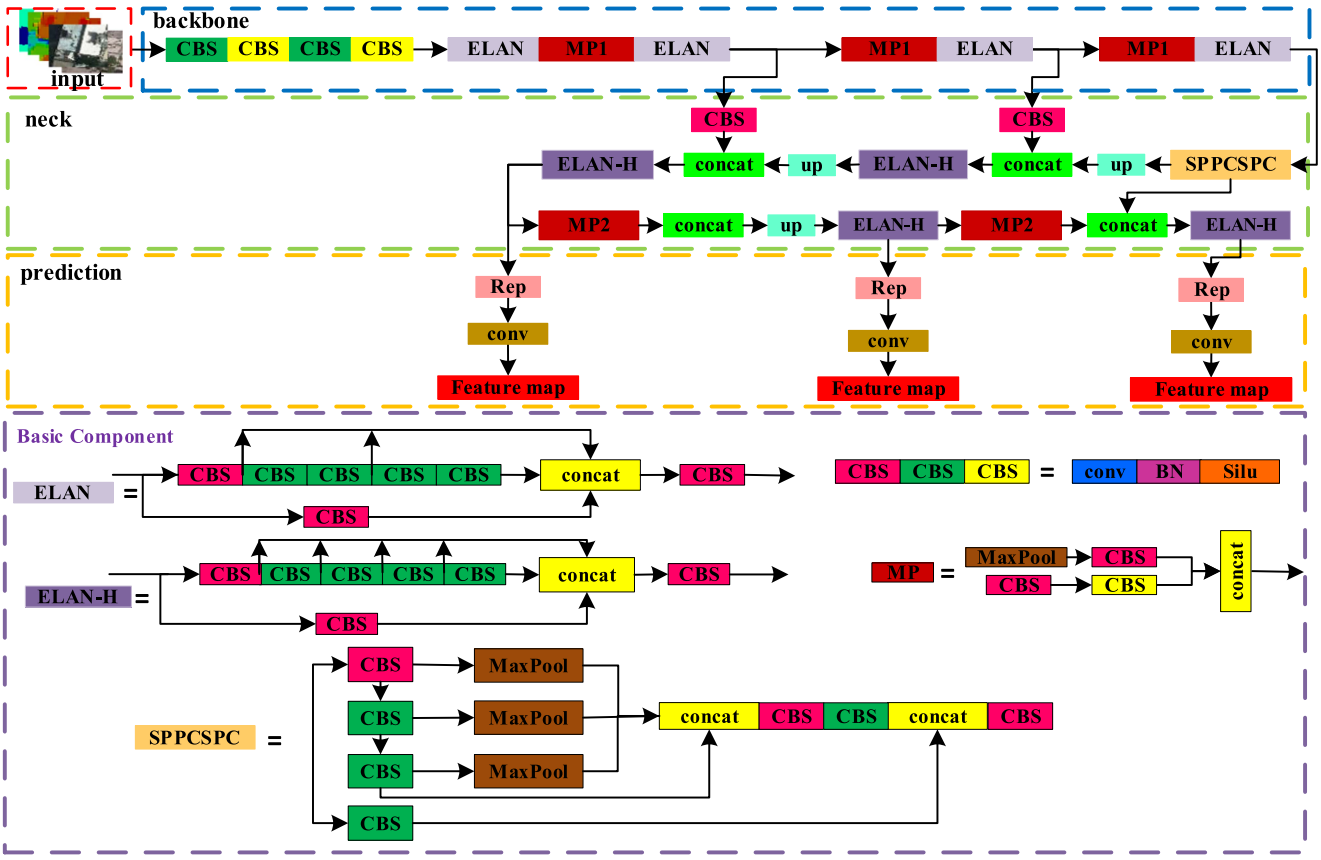


Fig. 5. Network architecture diagram of Yolov7 consists of three parts: backbone, neck, and head, and five basic components: CBS; MPconv; ELAN; ELAN-H; and SPPCSPC.

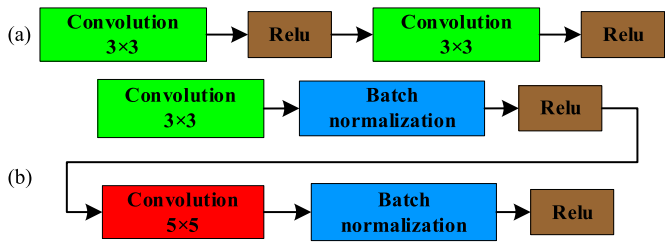


Fig. 6. (a) Convolutional layer of Unet and (b) the convolutional layer of MUnet.

Consequently, the encoder–decoder network of this article is made up of networks Yolov7 and CLMUnet.

IV. Experimental Results

In this section, we examine the developed method and compare the performances in the case of several encoder paths. This section includes Experimental parameter settings, Evaluation metrics and Comparison of experimental results.

A. Experimental Parameter Settings

The experimental environment is Intel(R) core (TM) i7-7800X CPU 3.5GHz, 32.0GB installed RAM, NVIDIA GeForce GTX 1050Ti, and all studied deep learning networks are trained using Tensorflow 2.10.0 and Python 3.8. The datasets are partitioned into patches with dimensions of 128×128 due to the limitation of running memory. Initially 600 patches obtained from the first dataset and the MA step adds up 312 patches. 60% of these 912 patches considered as the train data and the remaining 40% are used for the test data. After "train test split" the augmentation (+90° and +180° rotation) is applied to the train data so that it raises from 547 to 1641. For the second dataset 266 patches obtain initially and MA step increases them to 316. As before they split into 60% train and 40% test data sets. The augmentation (+90°, −90°, and +180° rotation) increases the train data to 756 patches. The focal loss function is usually used to classify unbalanced data for binary classification [58]. Since the map of final changes in this article is multiple, the focal loss function is expressed as follows [68]:

$$L_{fl} = -\sum_{i=1}^{c} \alpha_i (1-y_i)^{\gamma} t_i \log(y_i). \tag{3}$$

In this equation, C is the number of classes, t_i represents the real probability distribution, y_i represents the probability distribution of the prediction and γ and α_i are hyperparameters related to the focal loss function. Table II gives the other parameters applied to train the network in this article.

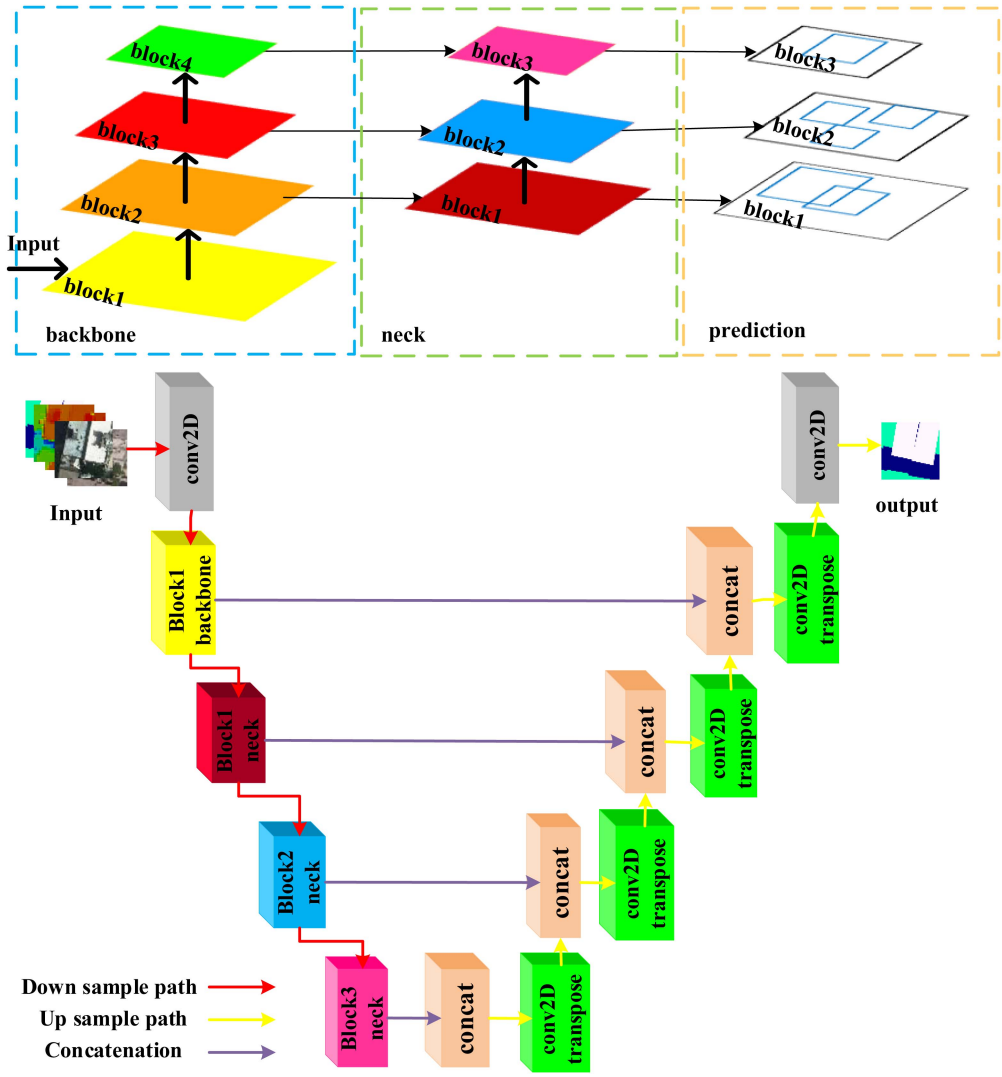


Fig. 7. Top: The structure of Yolov7 which contains three main parts: backbone, neck, and prediction (each block of neck and backbone represented with a different color); bottom: the structure of the proposed encoder–decoder network using Yolov7 as encoder and CLMUnet as a decoder. The encoder is concatenated with the decoder at four different resolutions (Block1 backbone, Block1 neck, Block2 neck, and Block3 neck).

TABLE II
PARAMETERS USED FOR TRAINING PROPOSED METHOD

Parameters	Value
optimizer	Adam
Learning rate	1×10^{-4}
Learning decay	1×10^{-6}
Pre-trained (encoder path Yolov7)	MS COCO dataset
γ (Focal loss)	2
Number of epochs	90
α_i(Focal loss)	Class weights after MA

TABLE III
INFORMATION FORMULAS FOR ACCURACY ASSESSMENT METRICS

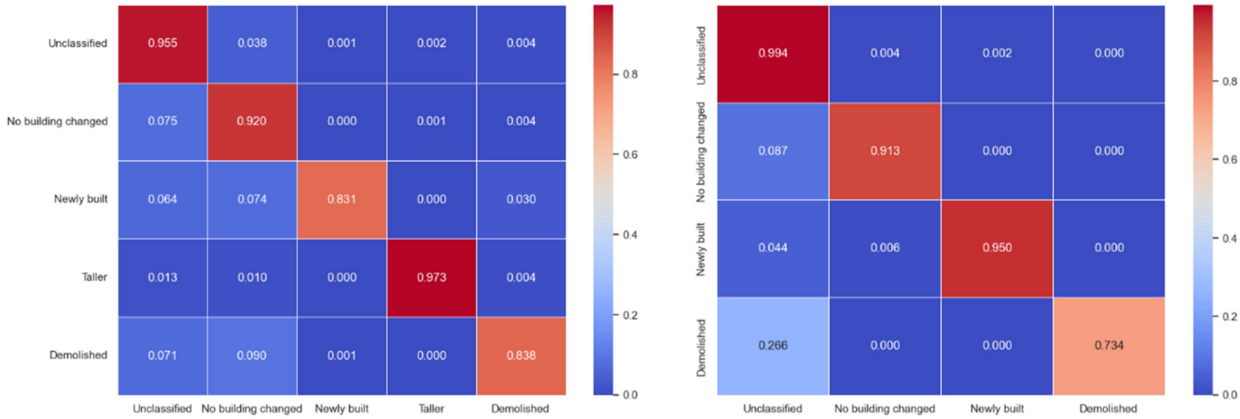
Evaluation metrics	Formula
Precision	$\frac{TP}{TP+FP}$
F1-score	$\frac{2\times TP}{(2\times TP)+FP+FN}$
IOU	$\frac{TP}{TP+FP+FN}$
Overall Accuracy (OA)	$\frac{TP+TN}{TP+FN+TN+FP}$
Kappa Coefficient (KC)	$\frac{2\times(TP\times TN-FN\times FP)}{(TP+FP)(FP+TN)+(TP+FN)(FN+TN)}$

B. Evaluation Metrics

Five evaluation metrics are stated in this article to evaluate the performance of the methods. *TP* indicates true positive (number of images predicted to be changed that were actually changed), *FP* denotes false positive (number of images predicted to be changed that were actually unchanged), *FN* denotes false

TABLE IV
QUANTITATIVE EVALUATION OF THE RESULTS

Encoder Path	OA (%)	Precision (%)	F1-Score (%)	IOU (%)	(KC)	Time of training (h min sec)	Parameters (Million)
Eff V2 B0	93.02	92.85	94.55	89.66	0.86	1 h 16 min 30 s	6.7
Eff V2 B1	93.12	94.56	94.76	90.00	0.86	1 h 25 min 30 s	7.1
Eff V2 B2	92.80	95.67	94.61	89.78	0.85	1 h 27 min 0 s	7.6
Eff V2 B3	92.77	94.49	94.46	89.50	0.86	1 h 33 min 44 s	8.8
Eff V2 T	93.08	94.30	94.72	89.94	0.86	1 h 47 min 23 s	8.9
Eff formerV2 L	93.37	93.32	94.86	90.22	0.87	1 h 56 min 20 s	6.7
Eff formerV2 S0	92.64	92.18	94.24	89.10	0.85	1 h 11 min 1 s	4.1
Eff formerV2 S1	92.43	92.47	94.13	88.90	0.85	1 h 23 min 10 s	4.6
Eff formerV2 S2	93.12	93.09	94.66	89.87	0.87	1 h 35 min 15 s	5.3
Tiny ViT	93.06	92.59	94.58	89.72	0.87	2 h 0 min 56 s	7.4
Yolox	88.64	92.75	91.49	84.32	0.75	1 h 3 min 28 s	4.9
Proposed method Mashhad UAV	94.87	96.14	96.08	92.46	0.89	1 h 29 min 32 s	12.3
Proposed method Tehran GeoEye-1	98.95	99.63	99.43	98.86	0.93	60 epochs 0 h 30 min 0 s	12.3



(a) (b)

Fig. 8. Confusion matrix of: (a) UAV of Mashhad dataset and (b) GeoEye-1 satellite stereo images of Tehran dataset.

negative (number of images predicted to be unchanged that were actually changed) and *TN* denotes true negative (number of images predicted to be unchanged that were actually unchanged) (Table III).

C. Comparison of Experimental Results

In Table IV, we used the evaluation metrics described in Section IV-B to compare the performance of several encoder paths.

We attempted to select versions of the networks that can be implemented in our system with input dimensions of 128×128. To achieve high accuracy at a reasonable speed and have the ability to manipulate the complexity of 3-D multiple CDs of buildings, the training time and the number of parameters of each network has conveniently considered in Table IV.

Although networks, such as EfficienNetV2 B0, B1, B2, Yolox, EfficienformerS0, and S2 have fewer parameters and faster training time, they could not handle the complexities of 3-D multiple CDs in the studied areas. The proposed network comprises an efficient number of parameters with high accuracy of the 3-D CD at a reasonable speed. It has the most parameters (12M) indicates that it can handle the necessary complexity for 3-D multiple CDs in buildings with suitable training time and achieves better results than the others. The Mashhad UAV dataset has the highest OA and KC of 94.87% and 0.9, respectively. In

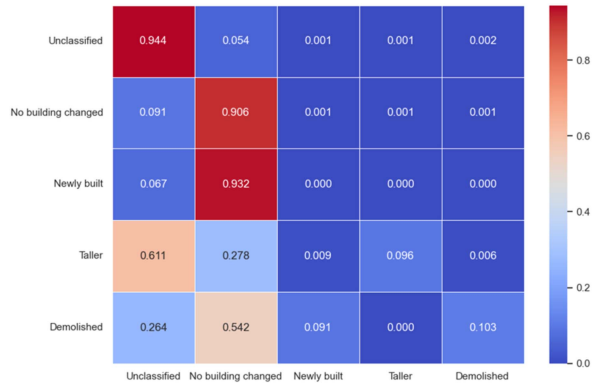


Fig. 9. Confusion matrix: UAV of Mashhad dataset without MA.

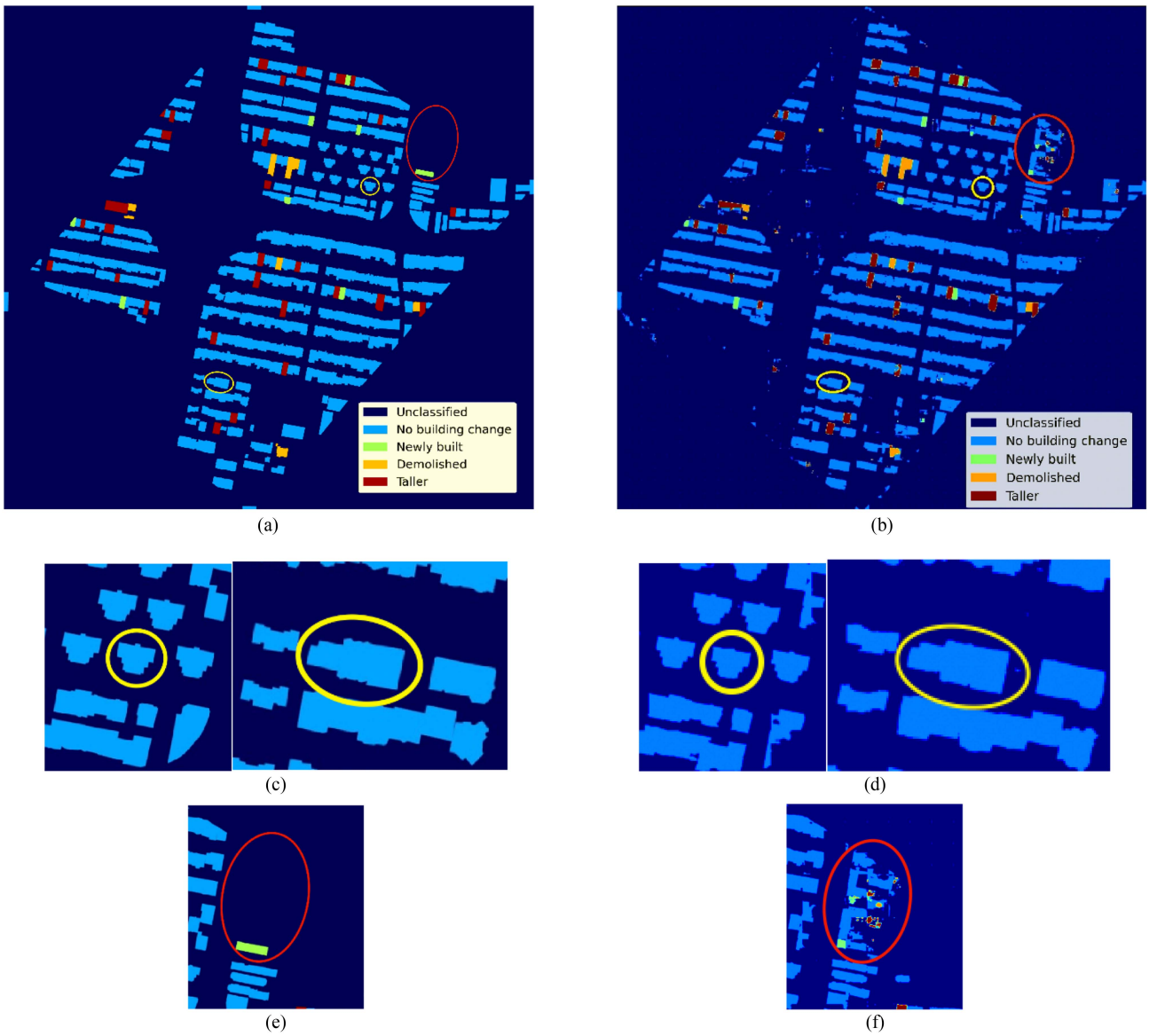


Fig. 10. Proposed method results for UAV of Mashhad dataset. (a) Ground truth. (b) Proposed method result, (c) two samples of buildings with different shapes marked with yellow ellipsoids. (d) Prediction of the two buildings by the proposed method. (e) Walled building-free area in ground truth, and (f) the corresponding prediction result of the walled area.

terms of other metrics, this network has achieved the best results when compared to other networks, demonstrates the method's proper performance.

Table IV gives the performance of proposed network on the GeoEye-1 Tehran dataset which after 60 epochs it achieved OA and KC of 98.95% and 0.93, respectively, indicates its appropriate performance on different datasets with variant complexities.

Fig. 8 depicts the confusion matrix associated with the network results to assess the performance of the proposed method detecting each class. The Mashhad UAV dataset has five classes: *Unclassified* which is related to urban objects other than the buildings; *no building change* which is related to the class of buildings with no change; *newly built* which is related to the class of buildings that did not exist in the first time and were built in the second time; *taller* which is related to the class of buildings that existed in the first time, but in the second time a newer building has taken its place with a higher height and finally; and *demolished* which is related to the class of buildings that are built in this place in the first time, but the building is destroyed in the second time. According to Fig. 8(a), the two classes with the fewest pixels (i.e., *newly built* and *demolished*) are 83% correctly recognized and the other classes are more than 92% correctly recognized. The GeoEye-1 dataset of Tehran includes four classes: *unclassified*; *no building change*; *newly built;* and *demolished* which is an urban developing area. The amount of the *demolished* class is much less than the other classes and is obtained with reasonably good accuracy. The majority of building changes related to construction and urban development have been detected 95% correctly in the confusion matrix.

In [50], an object-based framework is applied to our second dataset and the result is compared with two other methods. The comparison of some evaluation metrics is given in Table V.

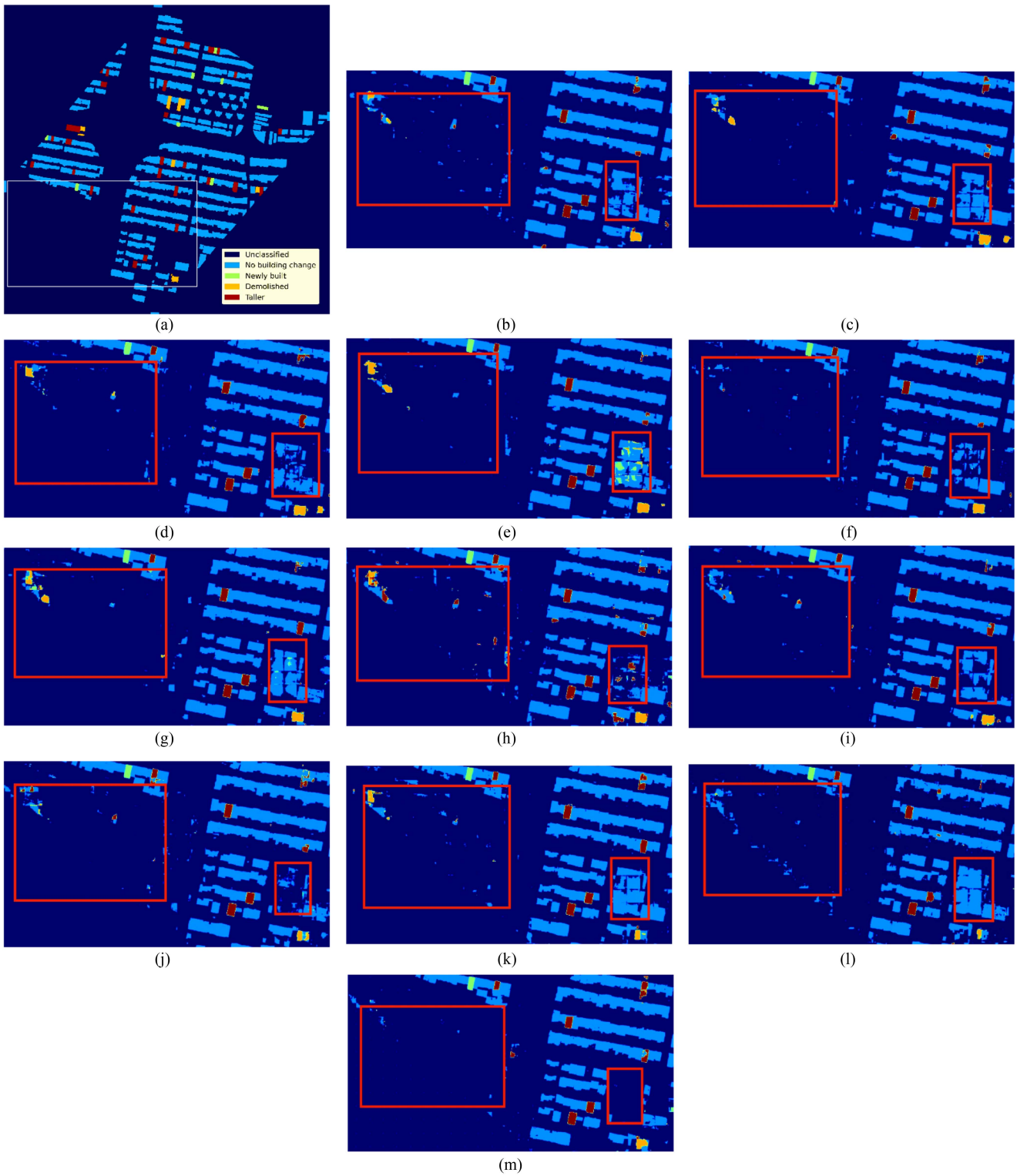


Fig. 11. Comparison between the various network selections as encoder path. (a) Ground truth in which the white rectangle indicates the area of networks results. (b) Eff V2 B0. (c) Eff V2 B1. (d) Eff V2 B2. (e) Eff V2 B3. (f) Eff V2 T. (g) Eff formerV2 L. (h) Eff formerV2 S0. (i) Eff formerV2 S1. (j) Eff formerV2 S2. (k) Tiny ViT. (l) Yolox. (m) Proposed method.

D. Ablation Study

One of our strengths in the proposed method is utilizing mixed-augmentation technique in 3-D CD method. Table VI gives the influence of MA step as an ablation study in the first dataset CD. One can observe that the proposed method without MA step has lower performance. Furthermore, the confusion matrix in Fig. 9 emphasizes the impact of MA specially in detecting *newly built*, *taller* and *demolished* classes in comparison with Fig. 8(a).

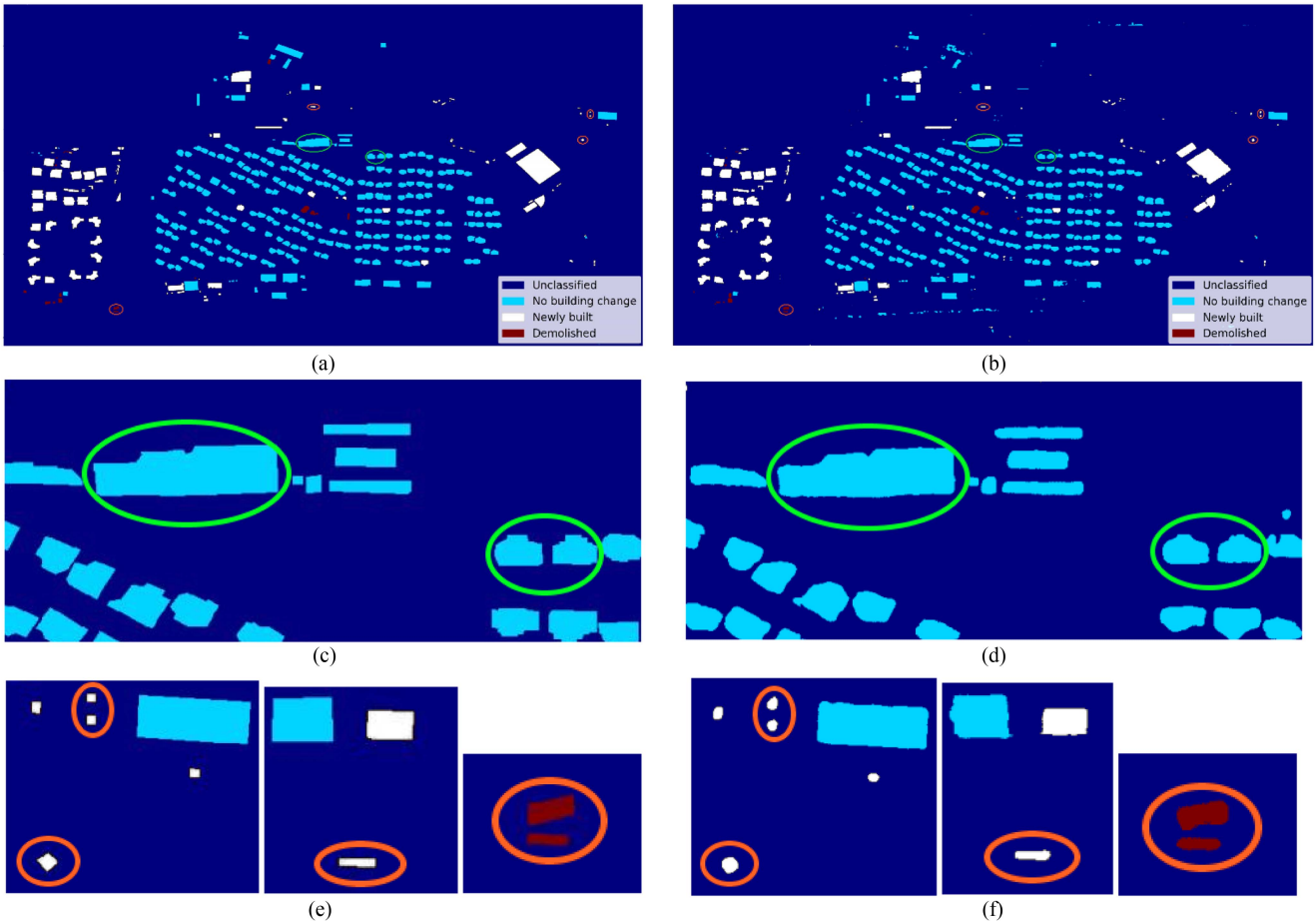


Fig. 12. Proposed method results for GeoEye-1 satellite stereo images of Tehran dataset. (a) Ground truth. (b) proposed method result. (c) Two samples of buildings with different shapes marked with green ellipsoids. (d) Prediction of the two buildings by the proposed method. (e) Four samples of small changes. (f) Prediction of the small changes by the proposed method.

TABLE V
QUANTITATIVE COMPARISON OF THE SECOND DATASET RESULTS

Method	OA (%)	F1 Score (%)
DSM differencing [69]	69.6	81.7
Postrefinement [70]	89.1	94.1
Objectbased method [50]	97.9	90.7
Proposed method	98.9	99.4

TABLE VI
ABLATION STUDY OF THE MA IN THE CASE OF THE FIRST DATASET

categories	OA (%)	Precision (%)	F1-Score (%)	IOU (%)	KC
With MA	94.87	96.14	96.08	92.46	0.89
Without MA	90.07	94.62	94.22	89.08	0.80

V. DISCUSSION

The performance of the proposed method in 3-D multiple CD in buildings is assessed by quantitative and qualitative evaluation as well as visual analysis. The proposed method's results are compared to the ground truth in Fig. 10. The first data set includes five classes: *unclassified* (dark blue); *no building change* (light blue); *newly built* (light green); *demolished* (orange); and *taller* (crimson). As shown in Fig. 10, the *newly built* and *demolished* classes have the lower frequencies which is one of the major challenges of this research and is overcome by employing the MA method. The area of Mashhad City has a lot of diversity and urban complexity, such as the discrepancy in geometric shapes of the buildings and spectral and texture differences of their roofs make 3-D CD difficult, so the proposed method should be able to manipulate this variety and finally provide a map of multiple changes with high accuracy. Yellow circles mark two buildings with completely different geometric shapes in Fig. 10(c) and (d). Comparing the network results with the ground truth indicates that the proposed method can correctly recognize the shape of buildings as well as preserve their edges. Furthermore, while the roofs of buildings have sometimes the same color as other urban features, such as streets or roads, the proposed method can correctly detect the buildings. The red circle in Fig. 10(e) and (f) marks a walled building-free area which is difficult to distinguish from the buildings, even by the human vision, since the spectral and texture of this area are very close to the buildings and also have the same height as buildings due to the walls.

Fig. 11 compares the performance between the various network selections as encoder paths along with a white rectangle of the ground truth. Red rectangles mark an open-air stadium and an urban green area. Other networks could not function properly

and had significant errors in these areas, but the proposed method had the best performance and the least amount of error in the final multiple changes map.

The second studied area has four classes, as shown in Fig. 12(a): *unclassified* (dark blue); *no building change* (light blue); *newly built* (yellow); and *demolished* (crimson). In Fig. 12(c) and (d), two examples of buildings with different geometric shapes are marked by green ellipsoids which the proposed method successfully preserved and detected their geometric shape and the edge of the buildings. Furthermore, the proposed method was able to detect small changes marked by orange ellipsoids that was related to small buildings or commercial centers.

VI. Conclusion

In this article, we developed a network based on the encoder–decoder architecture for 3-D multiple building CDs. We examine more than 15 networks as an encoder path and among them, blocks of Yolov7 give the best performance and accuracy at reasonable speed. We utilized this network pre-trained by the MS COCO dataset as the encoder path to exploit the benefits of the semi-transfer learning technique, and employed the convolution layers of the MUnet network as the decoder path. This method is applied to two datasets with multimodality and complexity. The first dataset is the images and the point clouds taken by UAV in two epochs 2011 and 2016 from a densely built urban area in the city of Mashhad, Iran, which includes buildings with various geometric shapes, spectrum, and textural roof structures. The second dataset includes two stereo-images of the GeoEye-1 satellite of the city of Tehran, Iran, in epochs 2009 and 2013. The semi-global matching technique is employed to generate the point clouds from these stereo-images. The remarkable thing about these two datasets is that they are both highly unbalanced so the MA technique is applied to resolve this problem. The method investigated in this article achieved an OA of 94.81% and 98.95%, and also a KC of 0.89 and 0.93 for the first and second datasets, respectively. The results show that the constructed method was successful to overcome the geometric and height diversity of the buildings, as well as the spectral and textural diversity of their roofs. Furthermore, an examination of the confusion matrix reveals that the method investigated in this article is capable of accurately distinguishing each class.

According to the appropriate performance of the method studied in this article which can automatically detect 3-D multiple changes of buildings and cope with the complexities of the studied areas, it is suggested that this method can be applied to further 3-D multiple CD in other urban areas with different complexities which have different kinds of highly unbalanced datasets, such as Lidar and radar.

References

[1] T. Ku et al., "SHREC 2021: 3D point cloud change detection for street scenes," *Comput. Graph.*, vol. 99, pp. 192–200, 2021.

[2] D. P. Argialas, S. Michailidou, and A. Tzotsos, "Change detection of buildings in suburban areas from high resolution satellite data developed through object based image analysis," *Surv. Rev.*, vol. 45, no. 333, pp. 441–450, 2013.

[3] M. Bouziani, K. Goïta, and D. - C. He, "Automatic change detection of buildings in urban environment from very high spatial resolution images using existing geodatabase and prior knowledge," *ISPRS J. Photogramm. Remote Sens.*, vol. 65, no. 1, pp. 143–153, 2010.

[4] X. Huang, L. Zhang, and T. Zhu, "Building change detection from multitemporal high-resolution remotely sensed images based on a morphological building index," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 1, pp. 105–115, Jan. 2013.

[5] S. Pang, X. Hu, M. Zhang, Z. Cai, and F. Liu, "Co-segmentation and superpixel-based graph cuts for building change detection from bitemporal digital surface models and aerial images," *Remote Sens.*, vol. 11, no. 6, 2019, Art. no. 729.

[6] P. Li, H. Xu, and J. Guo, "Urban building damage detection from very high resolution imagery using OCSVM and spatial features," *Int. J. Remote Sens.*, vol. 31, no. 13, pp. 3393–3409, 2010.

[7] R. Qin, J. Tian, and P. Reinartz, "3D change detection–approaches and applications," *ISPRS J. Photogramm. Remote Sens.*, vol. 122, pp. 41–56, 2016.

[8] S. T. Seydi and M. Hasanlou, "A new structure for binary and multiple hyperspectral change detection based on spectral unmixing and convolutional neural network," *Measurement*, vol. 186, Dec. 2021, Art. no. 110137.

[9] A. Menderes, A. Erener, and G. Sarp, "Automatic detection of damaged buildings after earthquake hazard by using remote sensing and information technologies," *Procedia Earth Planet. Sci.*, vol. 15, pp. 257–262, 2015.

[10] A. Sasagawa, E. Baltsavias, S. Kocaman-Aksakal, and J. D. Wegner, "Investigation on automatic change detection using pixel-changes and DSM-changes with ALOS-PRISM triplet images," *Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 40, no. 7/W2, pp. 213–217, 2013.

[11] J. Tian, H. Chaabouni-Chouayakh, P. Reinartz, T. Krauss, and P. d'Angelo, "Automatic 3D change detection based on optical satellite stereo imagery," *Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 38, no. 7B, pp. 586–591, 2010.

[12] X. Tong et al., "Building-damage detection using pre-and post-seismic high-resolution satellite stereo imagery: A case study of the May 2008 Wenchuan earthquake," *ISPRS J. Photogramm. Remote Sens.*, vol. 68, pp. 13–27, 2012.

[13] N. Champion, D. Boldo, M. Pierrot-Deseilligny, and G. Stamon, "2D building change detection from high resolution satellite imagery: A two-step hierarchical method based on 3D invariant primitives," *Pattern Recognit. Lett.*, vol. 31, no. 10, pp. 1138–1147, 2010.

[14] W. Xiao, B. Vallet, and N. Paparoditis, "Change detection in 3d point clouds acquired by a mobile mapping system," *ISPRS Ann. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 1, no. 2, pp. 331–336, 2013.

[15] W. Boonpook, Y. Tan, H. Liu, B. Zhao, and L. He, "Uav-based 3D urban environment monitoring," *ISPRS Ann. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 4, no. 3, pp. 37–43, 2018.

[16] R. Qin, "Change detection on LOD 2 building models with very high resolution spaceborne stereo imagery," *ISPRS J. Photogramm. Remote Sens.*, vol. 96, pp. 179–192, 2014.

[17] H. Chaabouni-Chouayakh, P. d'Angelo, T. Krauss, and P. Reinartz, "Automatic urban area monitoring using digital surface models and shape features," in *Proc. Joint Urban Remote Sens. Event*, 2011, pp. 85–88.

[18] S. Pang, X. Hu, Z. Wang, and Y. Lu, "Object-based analysis of airborne LiDAR data for building change detection," *Remote Sens.*, vol. 6, no. 11, pp. 10733–10749, 2014.

[19] J. Rogan, J. Franklin, and D. A. Roberts, "A comparison of methods for monitoring multitemporal vegetation change using thematic mapper imagery," *Remote Sens. Environ.*, vol. 80, no. 1, pp. 143–156, 2002.

[20] C. Stal, F. Tack, P. De Maeyer, A. De Wulf, and R. Goossens, "Airborne photogrammetry and lidar for DSM extraction and 3D change detection over an urban area–a comparative study," *Int. J. Remote Sens.*, vol. 34, no. 4, pp. 1087–1110, 2013.

[21] J. Tian, L. Metzlaff, P. d'Angelo, and P. Reinartz, "Region-based building rooftop extraction and change detection," *Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 42, pp. 903–908, 2017.

[22] J. Tian, J. Dezert, and P. Reinartz, "Refined building change detection in satellite stereo imagery based on belief functions and reliabilities," in *Proc. IEEE Int. Conf. Multisensor Fusion Integration Intell. Syst.*, 2015, pp. 160–165.

[23] C. Dai, Z. Zhang, and D. Lin, "An object-based bidirectional method for integrated building extraction and change detection between multimodal point clouds," *Remote Sens.*, vol. 12, no. 10, 2020, Art. no. 1680.

[24] R. Qin, X. Huang, A. Gruen, and G. Schmitt, "Object-based 3-D building change detection on multitemporal stereo images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 5, pp. 2125–2137, 2015.

[25] D. Wen, X. Huang, A. Zhang, and X. Ke, "Monitoring 3D building change and urban redevelopment patterns in inner city areas of Chinese megacities using multi-view satellite imagery," *Remote Sens.*, vol. 11, no. 7, 2019, Art. no. 763.

[26] S. T. Seydi, M. Hasanlou, and J. Chanussot, "DSMNN-Net: A deep Siamese morphological neural network model for burned area mapping using multispectral sentinel-2 and hyperspectral PRISMA images," *Remote Sens.*, vol. 13, no. 24, Jan. 2021, Art. no. 24.

[27] M. Gomroki, M. Hasanlou, and P. Reinartz, "STCD-EffV2T Unet: Semi transfer learning EfficientNetV2 T-Unet network for urban/land cover change detection using sentinel-2 satellite images," *Remote Sens.*, vol. 15, no. 5, 2023, Art. no. 1232.

[28] X. Jiang, G. Li, X. - P. Zhang, and Y. He, "A semisupervised Siamese network for efficient change detection in heterogeneous remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, no. 1, pp. 1–18, Mar. 2021, doi: 10.1109/TGRS.2021.3061686.

[29] L. Mou and X. X. Zhu, "A recurrent convolutional neural network for land cover change detection in multispectral images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2018, pp. 4363–4366.

[30] S. T. Seydi, M. Hasanlou, and M. Amani, "A new end-to-end multi-dimensional CNN framework for land cover/land use change detection in multi-source remote sensing datasets," *Remote Sens.*, vol. 12, no. 12, 2020, Art. no. 2010.

[31] W. Zhang and X. Lu, "The spectral-spatial joint learning for change detection in multispectral imagery," *Remote Sens.*, vol. 11, no. 3, 2019, Art. no. 240.

[32] M. Zhang and W. Shi, "A feature difference convolutional neural network-based change detection method," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 10, pp. 7232–7246, Oct. 2020.

[33] Q. Ding, Z. Shao, X. Huang, and O. Altan, "DSA-Net: A novel deeply supervised attention-guided network for building change detection in high-resolution remote sensing images," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 105, 2021, Art. no. 102591.

[34] J. Ma, G. Shi, Y. Li, and Z. Zhao, "MAFF-Net: Multi-attention guided feature fusion network for change detection in remote sensing images," *Sensors*, vol. 22, no. 3, 2022, Art. no. 888.

[35] Y. Wang et al., "Mask DeepLab: End-to-end image segmentation for change detection in high-resolution remote sensing images," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 104, 2021, Art. no. 102582.

[36] Y. Zhang, M. Deng, F. He, Y. Guo, G. Sun, and J. Chen, "FODA: Building change detection in high-resolution remote sensing images based on feature–output space dual-alignment," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, no. 1, pp. 8125–8134, Aug. 2021, doi: 10.1109/JSTARS.2021.3103429.

[37] L. Moya et al., "Detecting urban changes using phase correlation and ℓ1-based sparse model for early disaster response: A case study of the 2018 Sulawesi Indonesia earthquake-tsunami," *Remote Sens. Environ.*, vol. 242, 2020, Art. no. 111743.

[38] Z. Zheng, Y. Zhong, J. Wang, A. Ma, and L. Zhang, "Building damage assessment for rapid disaster response with a deep object-based semantic change detection framework: From natural disasters to man-made disasters," *Remote Sens. Environ.*, vol. 265, 2021, Art. no. 112636.

[39] J. Chen, H. Liu, J. Hou, M. Yang, and M. Deng, "Improving building change detection in VHR remote sensing imagery by combining coarse location and co-segmentation," *ISPRS Int. J. Geo-Inf.*, vol. 7, no. 6, 2018, Art. no. 213.

[40] Z. Zhang, G. Vosselman, M. Gerke, D. Tuia, and M. Y. Yang, "Change detection between multimodal remote sensing data using siamese CNN," vol. 1, no. 1, pp. 1–17, Jul. 2018, *arXiv:1807.09562*.

[41] L. Bruzzone and S. B. Serpico, "An iterative technique for the detection of land-cover transitions in multitemporal remote-sensing images," *IEEE Trans. Geosci. remote Sens.*, vol. 35, no. 4, pp. 858–867, Jul. 1997.

[42] P. P. De Bem, O. A. de Carvalho Junior, R. Fontes Guimarães, and R. A. Trancoso Gomes, "Change detection of deforestation in the Brazilian Amazon using landsat data and convolutional neural networks," *Remote Sens.*, vol. 12, no. 6, 2020, Art. no. 901.

[43] R. Liu, M. Kuffer, and C. Persello, "The temporal dynamics of slums employing a CNN-based change detection approach," *Remote Sens.*, vol. 11, no. 23, 2019, Art. no. 2844.

[44] Z. Zhang, G. Vosselman, M. Gerke, C. Persello, D. Tuia, and M. Y. Yang, "Change detection between digital surface models from airborne laser scanning and dense image matching using convolutional neural networks," *ISPRS Ann. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 4, pp. 453–460, 2019.

[45] Z. J. Yew and G. H. Lee, "City-scale scene change detection using point clouds," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2021, pp. 13362–13369.

[46] R. Yadav, A. Nascetti, and Y. Ban, "Building change detection using multi-temporal airborne LiDAR data," Apr. 26, 2022, doi: 10.48550/arXiv.2204.12535.

[47] X. Lian, W. Yuan, Z. Guo, Z. Cai, X. Song, and R. Shibasaki, "End-to-end building change detection model in aerial imagery and digital surface model based on neural networks," *Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 43, pp. 1239–1246, 2020.

[48] B. Nagy, L. Kovács, and C. Benedek, "ChangeGAN: A deep network for change detection in coarsely registered point clouds," *IEEE Robot. Automat. Lett.*, vol. 6, no. 4, pp. 8277–8284, Jul. 2021.

[49] I. de Gélis, S. Lefèvre, and T. Corpetti, "3d urban change detection with point cloud Siamese networks," *Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 43, pp. 879–886, 2021.

[50] H. Mohammadi and F. Samadzadegan, "An object based framework for building change analysis using 2D and 3D information of high resolution satellite images," *Adv. Space Res.*, vol. 66, no. 6, pp. 1386–1404, 2020.

[51] H. Amini Amirkolaee and H. Arefi, "3D change detection in urban areas based on DCNN using a single image," *Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 42, pp. 89–95, 2019.

[52] C. - Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, pp. 7464–7475, 2022.

[53] H. Hirschmuller, "Stereo processing by semiglobal matching and mutual information," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 2, pp. 328–341, Feb. 2007.

[54] "Humpback whale identification | kaggle." Accessed: Nov. 04, 2023. [Online]. Available: https://www.kaggle.com/c/humpback-whale-identification

[55] G. King and L. Zeng, "Logistic regression in rare events data," *Political Anal.*, vol. 9, no. 2, pp. 137–163, 2001.

[56] K. C. Wong, M. Moradi, H. Tang, and T. Syeda-Mahmood, "3D segmentation with exponential logarithmic loss for highly unbalanced object sizes," in *Proc. 21st Int. Conf. Med. Image Comput. Comput. Assist. Intervention*, 2018, pp. 612–619.

[57] C. Summers and M. J. Dinneen, "Improved mixed-example data augmentation," in *Proc. IEEE Winter Conf. Appl. Comput. Vis.*, 2019, pp. 1262–1270.

[58] T. - Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2980–2988.

[59] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You look only once: Unified real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 779–788.

[60] I. Gallo, A. U. Rehman, R. H. Dehkordi, N. Landro, R. L. Grassa, and M. Boschetti, "Deep object detection of crop weeds: Performance of YOLOv7 on a real case dataset from UAV images," *Remote Sens.*, vol. 15, no. 2, 2023, Art. no. 539.

[61] Y. Wang, H. Wang, and Z. Xin, "Efficient detection model of steel strip surface defects based on YOLO-V7," *IEEE Access*, vol. 10, pp. 133936–133944, 2022.

[62] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervention*, 2015, pp. 234–241.

[63] D. Peng, Y. Zhang, and H. Guan, "End-to-end change detection for high resolution satellite images using improved UNet++," *Remote Sens.*, vol. 11, no. 11, 2019, Art. no. 1382.

[64] M. Tan and Q. Le, "Efficientnetv2: Smaller models and faster training," in *Proc. Int. Conf. Mach. Learn.*, 2021, pp. 10096–10106.

[65] Y. Li et al., "Efficientformer: Vision transformers at mobilenet speed," *Adv. Neural Inf. Process. Syst.*, vol. 35, pp. 12934–12949, 2022.

[66] K. Wu et al., "Tinyvit: Fast pretraining distillation for small vision transformers," in *Proc. 17th Eur. Conf. Comput. Vis.*, 2022, pp. 68–85.

[67] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "Yolox: Exceeding yolo series in 2021," Aug. 05, 2021, doi: 10.48550/arXiv.2107.08430.

[68] W. Liu, L. Chen, and Y. Chen, "Age classification using convolutional neural networks with the multi-class focal loss," in *Proc. IOP Conf. Ser., Mater. Sci. Eng.*, 2018, Art. no. 012043.

[69] H. Murakami, K. Nakagawa, H. Hasegawa, T. Shibata, and E. Iwanami, "Change detection of buildings using an airborne laser scanner," *ISPRS J. Photogramm. Remote Sens.*, vol. 54, no. 2–3, pp. 148–152, 1999.

[70] S. Pang, X. Hu, Z. Cai, J. Gong, and M. Zhang, "Building change detection from bi-temporal dense-matching point clouds and aerial images," *Sensors*, vol. 18, no. 4, 2018, Art. no. 966.

Masoomeh Gomroki received the B.Eng. degree in geomatics engineering, in 2012 and the M.Eng. degree in photogrammetry from the University of Isfahan, Iran, in 2015. She is currently working toward the Ph.D. degree in photogrammetry and remote sensing with Tehran University, Tehran, Iran.

Her research interests include the application of deep learning in remote sensing, change detection, medical image processing, language modeling, and LiDAR-based applications for forest analysis.

Mahdi Hasanlou (Member, IEEE) received the B.Sc. degree in surveying and geomatics engineering from the University of Tehran, Tehran, Iran, in 2003, the M.Sc. degree in remote sensing from the University of Tehran, Tehran, Iran, in 2006, and the Ph.D. degree in remote sensing from the University of Tehran, Tehran, Iran, in 2013.

Since 2013, he has been an Associate professor with the School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Iran, where he is the Head of the remote sensing laboratory. He is currently the Head of the remote sensing and photogrammetry group at this school. His research activities are mainly focused on hyperspectral, thermal, optical, and SAR remote sensing for urban and agro-environmental applications.

Jocelyn Chanussot (Fellow, IEEE) received the M.Sc. degree in electrical engineering from the Grenoble Institute of Technology (Grenoble INP), Grenoble, France, in 1995, and the Ph.D. degree in electrical engineering from the Universit de Savoie, Annecy, France, in 1998.

Since 1999, he has been with Grenoble INP, where he is currently a Professor of signal and image processing. He was a Visiting Scholar with Stanford University, Stanford, CA, USA, KTH, Stockholm, Sweden, and NUS, Singapore. Since 2013, he has been an Adjunct Professor with the University of Iceland, Reykjavik, Iceland, and the Chinese Academy of Sciences, Aerospace Information research Institute, Beijing, China. From 2015 to 2017, he was a Visiting Professor with the University of California, Los Angeles, CA, USA. His research interests include image analysis, hyperspectral remote sensing, data fusion, machine learning, and artificial intelligence.

Dr. Chanussot was the AXA Chair in remote sensing with the Chinese Academy of Sciences, Aerospace Information research Institute. He is the Founding President of IEEE Geoscience and Remote Sensing French Chapter (2007–2010), which received the 2010 IEEE GRS-S Chapter Excellence Award. He was the recipient of multiple outstanding paper awards. He was the Vice-President of the IEEE Geoscience and Remote Sensing Society, in charge of meetings and symposia (2017–2019). He was the General Chair of the first IEEE GRSSWorkshop on Hyperspectral Image and Signal Processing, Evolution in Remote Sensing. He was the Chair (2009–2011) and Co-Chair of the GRS Data Fusion Technical Committee (2005–2008). He was a Member of the Machine Learning for Signal Processing Technical Committee of the IEEE Signal Processing Society (2006–2008) and the Program Chair of the IEEE InternationalWorkshop on Machine Learning for Signal Processing (2009). He is currently an Associate Editor for IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE TRANSACTIONS ON IMAGE PROCESSING, and PROCEEDINGS OF THE IEEE. He was the Editor-In-Chief for IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS and Remote Sensing (2011–2015). In 2014, he was a Guest Editor for IEEE SIGNAL PROCESSING MAGAZINE. He is a Member of the Institut Universitaire de France (2012–2017) and a Highly Cited Researcher (Clarivate Analytics/Thomson Reuters).