

3D CHANGE DETECTION IN URBAN AREAS BASED ON DCNN USING A SINGLE IMAGE

H. Amini Amirkolaei¹, H. Arefi^{1*}

¹ School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran
(hamed.amini, hossein.arefi) @ut.ac.ir

Commission VI, WG VI/4

KEY WORDS: Change detection, LiDAR, Single image, CNN, DSM, Urban area

ABSTRACT:

In this paper, a novel approach is proposed for 3D change detection in urban areas using only a single satellite images. To this purpose, a dense convolutional neural network (DCNN) is utilized in order to estimate a digital surface model (DSM) from a single image. In this regard, a densely connected convolutional network is employed for feature extraction and an upsampling method based on dilated convolution is employed for estimating the height values. The proposed DCNN is trained using satellite and Light Detection and Ranging (LiDAR) data which are provided in 2012 from Isfahan, Iran. Subsequently, the trained network is utilized in order to estimate DSM of a single satellite image that is provided in 2006. Finally, the changed areas are detected by subtracting the estimated DSMs. Evaluating the accuracy of the detected changed areas indicates 66.59, 72.90 and 67.90 for correctness, completeness, and kappa, respectively.

1. INTRODUCTION

Change detection and processing of the multi-temporal images is an important and challenging topic in the last decades and has many application in various fields such as urban planning, environmental monitoring, and disaster management (Lu et al., 2004). Nowadays, remote sensing data over the urban scenes can be provided by airborne space-borne imaging, Light Detection and Ranging (LiDAR), Synthetic Aperture Radar (SAR), etc. (Zhang et al., 2018). There are many studies in change detection and various techniques are developed in this topic. The conventional methods employed manual photo analysis techniques for detecting changes by comparing the images that are captured on different dates. Most of the techniques which are applied to the digital aerial and satellite images mainly utilized the radiometric differences between images for detecting the changed areas (Jung, 2004). Only using radiometric characteristics is not sufficient for obtaining accurate results (Waser et al., 2007). Height data is a valuable source of information for change detection which can be provided by dense matching and LiDAR. Change detection requires height information of different dates. Turker and Cetinkaya (2005), generated DSMs for both pre- and post-earthquake data and identified the collapsed buildings by computing the difference between generated DSMs (Turker and Cetinkaya, 2005). The capability of the LiDAR system makes it possible to acquire a 3D geospatial information with high accuracy from the Earth's surface. Some researchers that used LiDAR data for change detection, compare the 3D information of LiDAR data with the existing 3D data such as raster maps, CAD models, and Digital Surface Model (DSM) of stereo matching. Vosselman (2004) detected the buildings using the LiDAR data and compared the results with the existing maps in for map updating (Matikainen et al., 2003; Vosselman et al.,

2004). Choi et al. (2009), presented a feature-based method for change detection in urban areas using the LiDAR data. They classified LiDAR data into pre-defined classes and determined the types of changes based on the classes (Choi et al., 2009). Wang and Li (2015) utilized high resolution satellite images beside LiDAR data in order to extract the collapsed buildings after an earthquake by combining the textural and radiometrical information of images by the 3D information of LiDAR data (Wang and Li, 2015). Generally, change detection is a complicated process and there is no optimal approach for all cases.

The robustness of Convolutional Neural Networks (CNN) in extracting deep and invariant features make them suitable for various topics in remote sensing such as object detection (Dong et al., 2015), height estimation (Amirkolaei and Arefi, 2019b; Ghamisi and Yokoya, 2018), as well as scene classification (Chen et al., 2016). Zhang et al. (2018) presented a framework in order to detect the buildings and trees changes using LiDAR data and aerial images. They employed a CNN to detect the candidate changes between two epochs and grouped the candidate changes into individual object changes (Zhang et al., 2018). Lim et al. (2018), proposed a CNN based algorithm for change detection from high resolution satellite images without any preprocessing, such as ortho-rectification and classification (Lim et al., 2018).

The main objective of this paper is to propose a dense convolutional neural network (DCNN) for DSM estimation from single remote sensing images in order to be employed further for 3D change detection analysis. The proposed DCNN is trained using an optical satellite image and LiDAR data for the first epoch and the trained network is employed to estimate a DSM from a single satellite image in second epoch. The changed areas are detected by comparing the estimated DSMs. The remainder of this paper is organized as follow, at first, the proposed deep framework is described in Section 2. Then, the

* Corresponding author

experimental result and analysis are presented in section 3. In Section 4 the conclusions and the suggestions for future work are presented.

2. PROPOSED METHODOLOGIES

In this section, the proposed DCNN for DSM estimation from a single remote sensing image and then detecting the changed areas are explained in details.

2.1 DSM estimation

DSM estimation from a single image in absence of environmental assumptions is a technically ill-posed problem due to the inherent ambiguity of mapping an remote sensing image into height values (Mou and Zhu, 2018). The ability of CNNs in the extracting various effective features in different scales from the input image make it suitable for estimating DSM from a single image. Indeed, in addition to the spectral and geometrical features which are extracted during the training procedure, the relation between pixels in different scales is investigated that is very effective for reconstructing the 3D structure of the objects. The architecture of the proposed DCNN is presented in Figure 1.

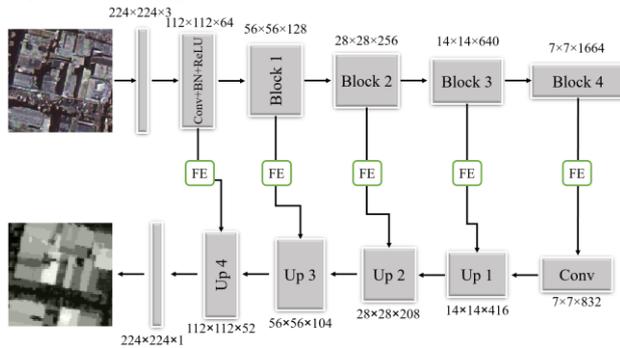


Figure 1. The proposed CNN architecture

The dense convolutional network (DenseNet) is a powerful CNN inspired by the residual network (ResNet) and is based on the repetition of a block to extract different and robust features from an optic image. Each layer of DenseNet is connected to all other layers in a feed-forward procedure in order to alleviate vanishing gradient, strengthen feature propagation and encourage feature reuse (Huang et al., 2017). In other words, instead of stacking L convolutional layers with L connections, the dense block has $\frac{L(L+1)}{2}$ connections and for each layer,

the feature maps of all preceding layers are utilized as inputs. In dense blocks, a 1×1 convolution is utilized as bottleneck before each 3×3 convolution in order to reduce the number of feature maps and improve the performance (Gómez-Ríos et al., 2019). The output of the i^{th} layer in the DenseNet is computed as follow:

$$l_i = H(C(l_{i-1}, l_{i-2}, \dots, l_0)) \quad (1)$$

where C is the concatenation operation, l is feature map, and H is a series of batch normalization, ReLU layer, and convolutional layer. The employed DenseNet contains four dense blocks that successively decrease the size of feature maps. It should be noted that the fully connected layer is removed for decreasing the number of parameters. Fig. 2 presents the structure of a dense block.

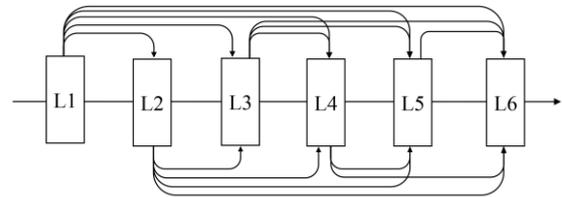


Figure 2. The structure of a dense block

In the DSM estimation, recovering the geometry structure of the objects is an important and challenging issue. As the generated feature maps after passing the input image through the DenseNet is very small ($1/32$ of the input image), the available information for recovering the geometry structure of the objects and recognizing the objects locations is not enough. Accordingly, proposing an efficient and robust upsampling approach is very important in order to reconstruct the structure of a region. Figure 3 represents the employed upsampling approach (Up in Figure 1) for DSM estimation (Laina et al., 2016).

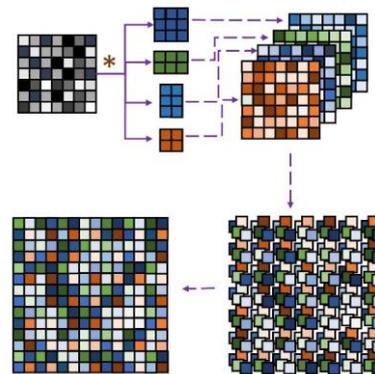


Figure 3. The structure of upsampling approach

In the utilized upsampling structure, four convolutional layers with 2×2 , 2×3 , 3×2 and 3×3 kernel size are applied into the input feature maps. It is obvious that the global structure of a considered region can be reconstructed using the generated coarse abstract feature maps after passing the contractive part. However, recovering local details and improving estimated height values would be possible only by considering the low-level features of shallow layers. Accordingly, an approach should be utilized in the upsampling approach to reuse the low-level feature maps of shallow layers and fuse them by the obtained high-level feature map of deep layers. The structure of the utilized approach for extracting context information from the shallow layers (FE in Figure 1) of the network and utilizing them in the upsampling part is represented in Figure 4.

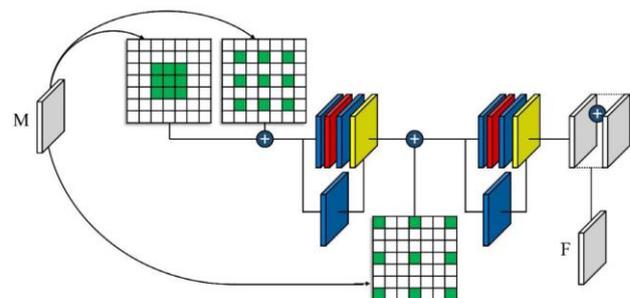


Figure 4. The structure of feature extraction from the shallow layers (Amirkolae and Arefi, 2019a).

In Figure 4, M is the feature maps of a shallow layer and F is the considering feature maps for upsampling. The context information depends on the receptive field of convolutional layers. A larger receptive field leads to greater context information to be captured. The basic method to increase the receptive field is to use the pooling operator which decreases the resolution of features maps. In order to alleviate this problem, dilated convolution is employed (Figure 4) that can increase the size of the receptive field without decreasing the size of the generated feature maps (Yu and Koltun, 2015). Let $D : Z^2 \rightarrow R$, $\Omega_r = [-r, r]^2 \cap Z^2$ and $k : \Omega_r \rightarrow R$ be a discrete of size $(2r + 1)^2$. The generalized form of the dilated convolution ($*$) is defined as follows (Yu and Koltun, 2015):

$$(D *_{l} k)(p) = \sum_{s+l=p} D(s)k(t), \quad (2)$$

The loss function plays an important role in CNNs to obtain an accurate result, making a determination of a suitable loss function is critical. The utilized loss function is consist of two main components. The first component is L2 norm (e_{L2}) that can eliminate the height shift between estimated DSM and reference DSM. The second component is slop value (e_s) that is computed using the adjacent pixels of each pixel and is suitable for improving the performance, especially, at the height jumps.

$$e_{L2} = \frac{1}{n} \sum_{i=1}^n (D_e - D_r)^2 \quad (3)$$

$$e_s = \frac{1}{n} \sum_{i=1}^n \sqrt{\left(\frac{\partial D_e}{\partial x} - \frac{\partial D_r}{\partial x}\right)^2 + \left(\frac{\partial D_e}{\partial y} - \frac{\partial D_r}{\partial y}\right)^2} \quad (4)$$

$$e = e_{L2} + e_s, \quad (5)$$

where h_r is the reference height value and h_e is the estimated height value.

2.2 Change detection

The proposed DCNN architecture in Section 2.1 is trained using single satellite image and LiDAR data of the first epoch. Afterward, the trained network is employed to estimate the DSM of a single image that belongs to another epoch. The changed areas are simply detected by subtracting the estimated DSMs of different dates. A greater similarity of the structures in the data makes the obtained result of subtracting contain more valuable and accurate information. Hence, the trained network is utilized for DSM estimation of both epochs. Subsequently, the estimated DSM of the first date is used instead of the LiDAR data for computing the difference image. Although the LiDAR data is more accurate than the estimated DSM, the similarity structure of the estimated DSMs, especially at boundaries of buildings, leads to reducing the noisy data and obtaining a better result.

3. EXPERIMENTS

In this section, in the beginning, the utilized dataset is described in Section 3.1. Then, the proposed approaches for DSM estimation (Section 3.1) and change detection (Section 3.2) are implemented.

3.1 Dataset

The utilized dataset consists of two remote sensing images and LiDAR data that are captured in 2006 and 2012 from Isfahan,

Iran. The provided data in 2012 consist of a pan-sharpened Pleiades image with 0.5 meter resolution and a LiDAR data which is resampled to a pixel size of 0.5 meter. The provided data in 2006 only contains a pan-sharpened QuickBird image with 0.5 meter resolution. The satellite images and LiDAR data of the study area are represented in Figure 5.

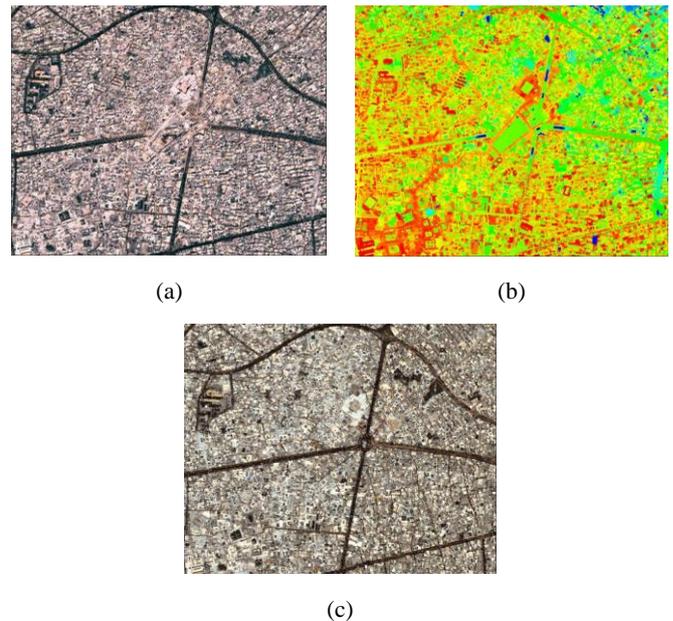
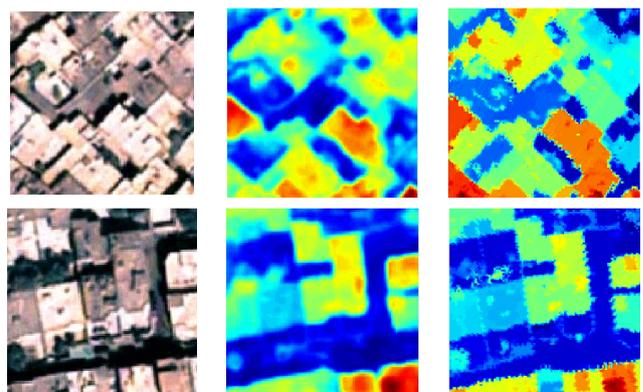


Figure 5. The study area, (a) Pleiades image in 2012, (b) LiDAR data in 2012, and (c) QuickBird image in 2006.

3.2 DSM estimation

To analyze the performance of the proposed DSM estimation approach the LiDAR data and Pleiades image in 2012 are employed. In this regard, firstly, the input image and corresponding LiDAR data are divided into smaller images (224×224) by considering to Figure 1. Almost 66 percent of the study area is selected for training and the remained 34 percent is selected for testing. Subsequently, the training data are augmented by applying certain transformation functions (Amirkolae and Arefi, 2019b). The number of training data after applying the transformations is approximately 77 k samples. The number of the epochs, batch size, learning rate, momentum rate, and weight decay are set to 15, 16, 10^{-3} , 0.9 and 3×10^{-4} , respectively. Some samples of the estimated DSMs are represented in Figure 6.



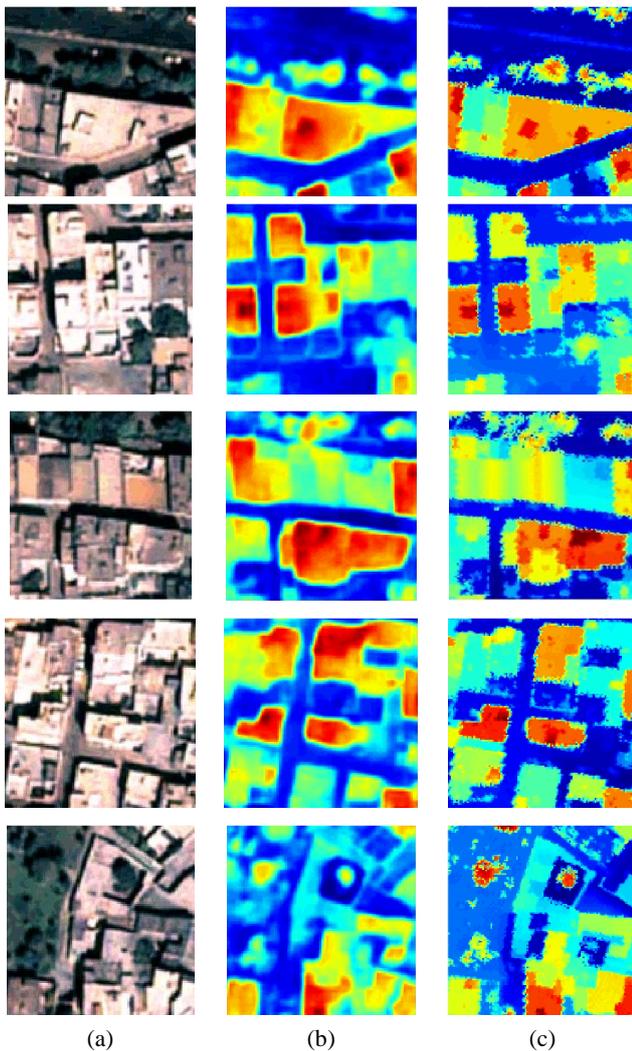


Figure 6. Some samples of the estimated DSMs, (a) input image, (b) estimated DSM, (c) reference DSM.

In order to compute the accuracy of the acquired results and assess the performance of the proposed DCNN, three criteria including root mean squared error (E_{RMSE}), average relative error (E_{Rel}), and average log10 error (E_{log10}) are computed (Amirkolaee and Arefi, 2019a).

$$E_{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (D_e - D_r)^2} \quad (6)$$

$$E_{rel} = \frac{1}{n} \sum_{i=1}^n \frac{|D_r - D_e|}{D_e} \quad (7)$$

$$E_{log10} = \frac{1}{n} \sum_{i=1}^n \left| \log_{10}^{D_r} - \log_{10}^{D_e} \right|^2 \quad (8)$$

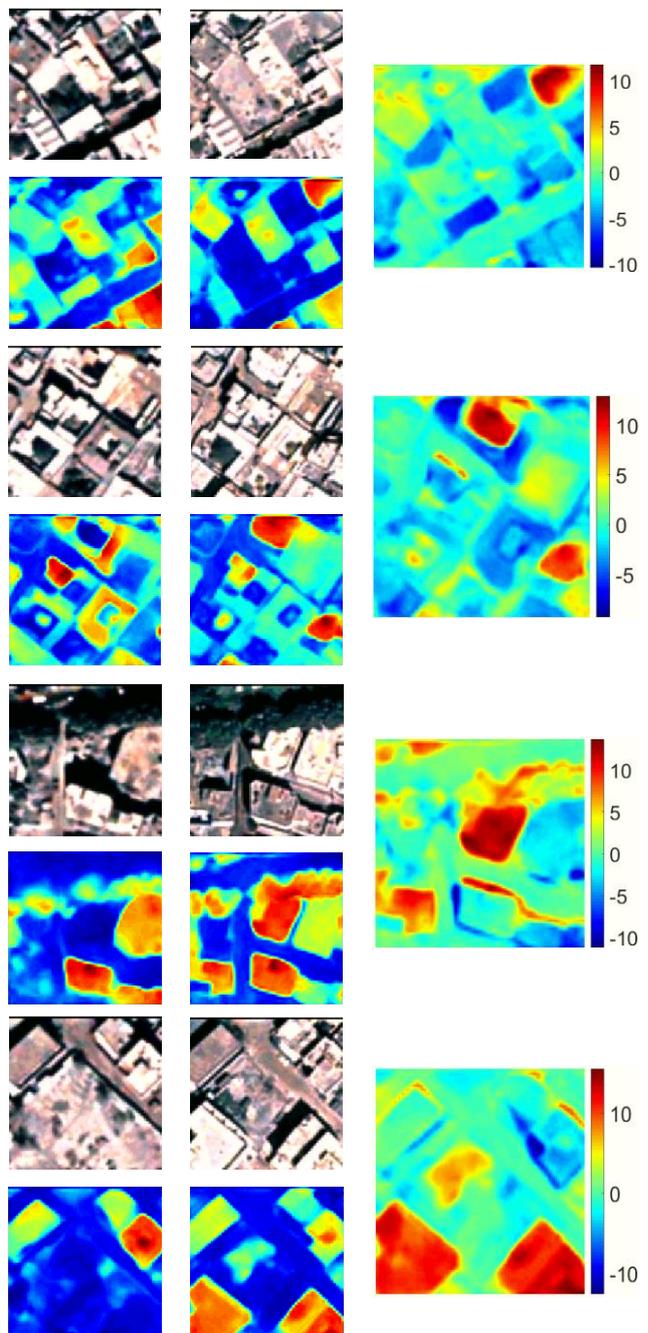
The E_{RMSE} is used to measure the absolute error at each pixel, the E_{Rel} is used to measure the ratio between absolute error and estimated height value and E_{log10} measures the absolute difference logarithm of the reference and estimated height values. The evaluation results of DSM estimation are represented in Table 1.

Table 1. The evaluation results of the DSM estimation

	E_{RMSE}	E_{Rel}	E_{log10}
Isfahan	2.487	0.731	0.274

3.3 Change detection

As mentioned in Section 2.2, the proposed DCNN is trained using the satellite image and LiDAR data which are captured in 2012 and the trained network is used to estimate the DSM of the satellite image that is provided in 2006. It should be noted, we also utilized the estimated DSM of 2012 based on DCNN approach instead of the LiDAR data to reduce noisy results. Figure 7 represents the results of change detection by subtracting the estimated DSMs. In the Fig 7d, the recognized changed areas after applying a threshold of 4 meter is represented. The positive changes which show the objects that are not exist in the 2006 are represented by red colour and the negative changes which show the object that are destroyed in the 2012 are represented by blue colour.



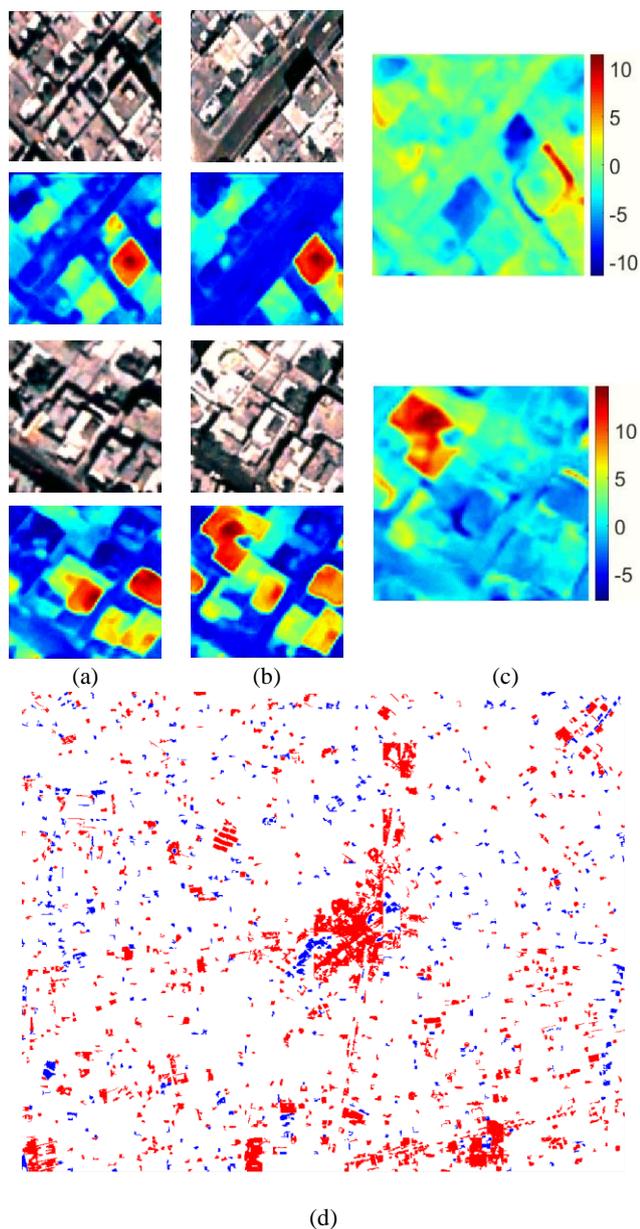


Figure 7. Some samples of the change detection, (a) the data of the 2006, (b) the data of the 2012, (c) the result of change detection, and (d) the positive and negative changed regions by applying a threshold (4 m) for the whole of the study area.

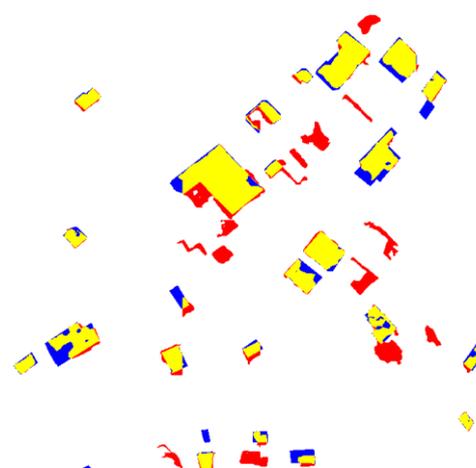
A reference data is manually produced in a small part of the study area. In order to assess the accuracy of the detected changed areas three criteria including, completeness (E_{Cm}), correctness (E_{Cr}) (Pahlavani et al., 2015), and Kappa (E_{Kappa}) (Niemeyer et al., 2013) are employed. The evaluation results of change detection is represented in Table 2 and Figure 8. The yellow area indicates the pixels that are correctly detected, the red pixels indicates the pixels which are wrongly detected and the blue pixels indicates the pixels which are not detected.

Table 2. The evaluation results of the change detection

	E_{Cr}	E_{Cm}	E_{Kappa}
Constructed	69.44	78.52	71.97
Destroyed	60.09	54.64	55.73
All	66.59	72.90	67.90



(a)



(b)

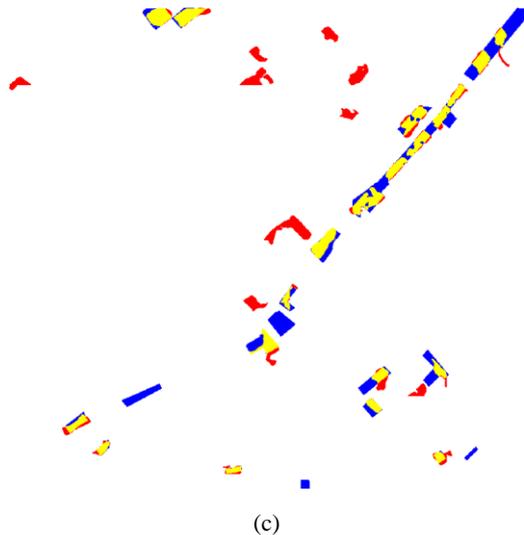


Figure 8. Schematic evaluation result, (a) reference data, (b) evaluation result for the constructed objects, (c) evaluation result for the destroyed objects.

4. CONCLUSION

In this paper, an approach is proposed for change detection from single remote sensing images. In this regard, a dense convolutional neural network (DCNN) is proposed for DSM estimation from a single remote sensing image. The network is trained using a satellite image and corresponding LiDAR data that are provided from Isfahan in 2012. Then, the trained network is used to estimate the DSM of a satellite image which is captured in 2006. The evaluation of the proposed DCNN for DSM estimation indicates root mean squared error, average relative error, and average log10 error of approximately 2.487 m, 0.731 and 0.274, respectively. In order to detect the changed areas, the estimated DSMs of 2012 is subtracted from the estimated DSM of 2006. The evaluation of the change detection results indicate 66.59, 72.90 and 67.90 for correctness, completeness, and kappa, respectively. Enriching the training data by employing the different images which are captured from the different satellites at different times can be suitable for improving the accuracy of the results that are suggested for future work.

REFERENCES

Amirkolaee, H.A., Arefi, H., 2019a. Convolutional neural network architecture for digital surface model estimation from single remote sensing image. *Journal of Applied Remote Sensing* 13, 016522.

Amirkolaee, H.A., Arefi, H., 2019b. Height estimation from single aerial images using a deep convolutional encoder-decoder network. *ISPRS Journal of Photogrammetry and Remote Sensing* 149, 50-66.

Chen, Y., Jiang, H., Li, C., Jia, X., Ghamisi, P., 2016. Deep feature extraction and classification of hyperspectral images based on convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing* 54, 6232-6251.

Choi, K., Lee, I., Kim, S., 2009. A feature based approach to automatic change detection from LiDAR data in urban areas. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci* 18, 259-264.

Dong, Z., Wu, Y., Pei, M., Jia, Y., 2015. Vehicle type classification using a semisupervised convolutional neural network. *IEEE transactions on intelligent transportation systems* 16, 2247-2256.

Ghamisi, P., Yokoya, N., 2018. Img2dsm: Height simulation from single imagery using conditional generative adversarial net. *IEEE Geoscience and Remote Sensing Letters* 15, 794-798.

Gómez-Ríos, A., Tabik, S., Luengo, J., Shihavuddin, A., Krawczyk, B., Herrera, F., 2019. Towards highly accurate coral texture images classification using deep convolutional neural networks and data augmentation. *Expert Systems with Applications* 118, 315-328.

Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q., 2017. Densely connected convolutional networks, *CVPR*, p. 3.

Jung, F., 2004. Detecting building changes from multitemporal aerial stereopairs. *ISPRS Journal of Photogrammetry and Remote Sensing* 58, 187-201.

Laina, I., Rupprecht, C., Belagiannis, V., Tombari, F., Navab, N., 2016. Deeper depth prediction with fully convolutional residual networks, 2016 Fourth international conference on 3D vision (3DV). *IEEE*, pp. 239-248.

Lim, K., Jin, D., Kim, C.-S., 2018. Change Detection in High Resolution Satellite Images Using an Ensemble of Convolutional Neural Networks, 2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC). *IEEE*, pp. 509-515.

Lu, D., Mausel, P., Brondizio, E., Moran, E., 2004. Change detection techniques. *International journal of remote sensing* 25, 2365-2401.

Matikainen, L., Hyypä, J., Hyypä, H., 2003. Automatic detection of buildings from laser scanner data for map updating. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 34, 218-224.

Mou, L., Zhu, X.X., 2018. IM2HEIGHT: Height estimation from single monocular imagery via fully residual convolutional-deconvolutional network. *arXiv preprint arXiv:1802.10249*.

Niemeyer, J., Rottensteiner, F., Soergel, U., 2013. Classification of urban LiDAR data using conditional random field and random forests, *Joint Urban Remote Sensing Event 2013*. *IEEE*, pp. 139-142.

Pahlavani, P., Amini Amirkolaee, H., Talebi Nahr, S., 2015. A new feature selection from lidar data and digital aerial images acquired for an urban/c environment using an ANFIS-based classification and a fuzzy rough set method. *International Journal of Remote Sensing* 36, 3587-3615.

Turker, M., Cetinkaya, B., 2005. Automatic detection of earthquake-damaged buildings using DEMs created from pre-and post-earthquake stereo aerial photographs. *International Journal of Remote Sensing* 26, 823-832.

Vosselman, G., Gorte, B., Sithole, G., 2004. Change detection for updating medium scale maps using laser altimetry. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 34, 207-212.

Wang, X., Li, P., 2015. Extraction of earthquake-induced collapsed buildings using very high-resolution imagery and airborne lidar data. *International Journal of Remote Sensing* 36, 2163-2183.

Waser, L., Baltsavias, E., Eisenbeiss, H., Ginzler, C., Grün, A., Kuechler, M., Thee, P., 2007. Change detection in mire ecosystems: assessing changes of forest area using airborne remote sensing data. *The international archives of the photogrammetry, remote sensing and spatial information sciences* 36, 313-318.

Yu, F., Koltun, V., 2015. Multi-scale context aggregation by dilated convolutions. *arXiv preprint arXiv:1511.07122*.

Zhang, Z., Vosselman, G., Gerke, M., Tuia, D., Yang, M.Y., 2018. Change detection between multimodal remote sensing data using siamese CNN. *arXiv preprint arXiv:1807.09562*.