



# Super Resolution Land Cover Mapping Using Deep Multi Scale Residual Dense Network

D. Synthiya Vinothini<sup>(✉)</sup>, B. Sathya Bama, Nirmal Selva, and Naveen Kumar

Thiagarajar College of Engineering, Madurai 625015, Tamilnadu, India  
{synthiya,sbece}@tce.edu, {nirmal,naveenkumar}@student.tce.edu

**Abstract.** Super Resolution Mapping (SRM) is a land cover mapping method that generates land surface cover maps at fine spatial resolution from a coarse spatial resolution Remote Sensing (RS) image. Recently deep networks have shown impressive performance for image super resolution and image segmentation. Inspired by this performance a deep Multi-Scale Residual Dense Network (MSRDN) is proposed for SRM application of satellite data which extracts hierarchical features that can efficiently map sub-pixels to an accurate class. A MSRDN network is trained with coarse resolution images and its corresponding fine resolution class cover patches to learn a super resolution mapping of land cover. The accuracy of the Conventional SRM techniques is restricted by the performance of soft classification methods. Hence this work utilizes the full power of deep learning to generate a fine resolution land cover map directly from a coarse resolution image neglecting the intermediate soft classification result. The results of the experiments show that MSRDN can become a best alternative to the conventional SRM techniques, to generate a precise land cover information directly from a coarse data.

**Keywords:** Super Resolution Mapping · Class proportion image · Coarse resolution · Fine resolution · Land cover class

## 1 Introduction

Whenever Land and Land cover is being spoken of, there is often a propensity to disregard the fact that it is a natural resource too, that needs preservation and wise usage. It is a limited resource we cannot afford to take lightly. Since most of the land cover is frequently being lost to floods, erosion, deforestation, overgrazing and other such adversities, effective Land cover monitoring is a subject of eternal importance, for its sustainable usage which requires accurate land cover mapping.

Accurate land cover mapping has important significance in multiple remote sensing applications like wetland inundation [1], urban floods map [2], city planning, environmental assessment and monitoring, land surface change detection, etc. But land cover mapping of remotely sensed imagery always suffers from a bottle neck of mixed-pixel problem especially when obtained from a coarse resolution sensor. Hard classification usually neglects this and consider only one class per pixel thus losing significant class

covers. Soft classification considers this mixed pixel problem and provides a convenient method to estimate the proportion of different classes within a mixed pixel. The core of Super Resolution Mapping (SRM) stems from the concept of dividing pixel into sub-pixels, thereby increasing the number of pixels per unit area providing a more useful High-Resolution Map for a coarse resolution image.

## 1.1 Related Works

Many SRM methods have been a proven method to address the mixed pixel problem. The most prevalent technique that addresses the mixed pixel problem is based on maximum spatial affinity model [3] in which a coarse mixed pixel is subdivided into subpixels. Each subpixel is assigned a class value such that it increases the spatial attractiveness within the pixel and its neighboring subpixels. This spatial affinity may be estimated at sub-pixel level [3], sub-pixel/pixel level [4, 5] or multiple levels [6, 7]. These spatially attractive models have been extensively used in SRM, but they are not appropriate for representing complex land cover structures like highly disjoint lands [8] also the model employed and the soft classifier used impacts the quality of the generated map strongly [9].

Many learning based SRM models have been developed to overcome the difficulty associated with explicit spatial dependency of land cover. These algorithms learn from an available fine resolution map. It works on the assumption that the coarse class proportion is related to the fine resolution map [8], thus the model learns this relationship from the available data to yield a SR map from a coarse class proportion image. These learning models can also be present as isolated ones or in fusion with another model. Learning models like back-propagation neural networks [10] and support vector regression [11] have been modelled to learn this relationship. Many hybrid learning models have also been developed to serve this purpose. Integration of Back Propagation Neural Network with Genetic Algorithm [1] for Super Resolution Mapping of Wetland Inundation wherein, a fusion of the GA and BPNN is present to ultimately perform SRM. The datasets used for this particular work were obtained from LANDSAT - TM/ETM Satellites, captured over two regions namely, Poyanghu in China and the Macquarie Marshes in Australia. This work was focused on Marshy Area mapping, and further works went on to explore the possibilities of Land Cover Mapping [12] as well as Urban Flood Mapping [2]. But practically its performance is limited in learning the complex non-linear relationships.

Currently in many computer vision and remote sensing applications Deep learning models have shown remarkable performance by learning the non-linear relationship between various parameters. One such work focused on Deep Learning for SRM is DeepSRM [13] that learns the non-linear relation between fraction classes and fine resolution map. But this class proportion is not readily available and not always reliable. Though SRM techniques yield enhanced class cover map with coarse class proportion image, its accuracy is always defined by the accuracy of the soft classification algorithm.

The implementation of Deep Learning in obtaining Super Resolution Maps (SRMap) is a domain which is still in the incubator, not many works have been reported in this regard, but a few, only in the recent past. Thus, a paradigm shift, leading to the development and application of various algorithms in SRM, became the primal subject of research. In this paper Deep Learning is implemented to learn a Multi-Scale Residual

Dense Network for SRM application to obtain High Resolution Maps directly from LR images as a single step process. Our work is devoted towards the induction of only Deep Learning in obtaining HRM, thereby excluding the possible use of class proportion data.

## 1.2 Contribution

Remarkable success in deep learning technology has led to the development of many computer vision task including Super resolution and image semantic segmentation to achieve greater heights. But very few networks have been reported for SRM application. Inspired by the performance of the Residual Dense Network (RDN) for image super resolution [14], this work adopts the RDN model to generate SRmap by exploiting the hierarchical features. Conventional SRM algorithms use class proportion data for mapping and its performance relies on the accuracy of the soft classification algorithm that gives the class proportion data. DeepSRM [13], the first paper on deep learning for SRM application also make use of this class proportion data. But the proposed work MSRDN, extracts multi scale features and does not rely on the accuracy of the output of soft classification but rather takes the full advantage of deep learning to yield an accurate land cover class map directly from a coarse image.

## 1.3 Organization

The rest of this paper is organized as follows. Section 2 introduces the proposed Multi-Scale Residual Dense Network for SRM. This section deals in detail about the network architecture and the process flow involved in training, testing and land cover generation. Section 3 presents the experimental results, quality measurements and its analysis. Finally, Sect. 4 concludes with a brief summary.

# 2 MSRDN: Multi-scale Residual Dense Network for SRM

Super Resolution Mapping generates a High-Resolution land cover information  $C_i$ , given a Low-Resolution image 'y', where C is a binary image representing the presence of  $i^{\text{th}}$  land cover class. The proposed model is trained to accurately generate the land cover information. This involves three steps: Dataset pre-processing, Training the MSRDN net, SRmap generation. Figure 1 show an overview of our proposed method for network training and SR land cover map generation.

## 2.1 Data Pre-processing

Generally Remote Sensing (RS) image data are too large to be processed by a network model in a single pass. The minimum dimension of the image data tile from the dataset Vaihingen provided by ISPRS is  $2336 \times 1281$ , which is way too large for a network's input considering the GPU memory limitations. The other constraint is the limited availability of RS data along with its ground truth, whereas a deep network requires more data for training. Considering both the GPU memory limitation and more training data

requirement the available large dataset tile and its corresponding ground truth information is divided into smaller overlapping patches with a simple sliding window. Now it is possible to process the entire large data linearly. The dataset consists of classified map for a remotely sensed image. The image is down-sampled to give a coarse image ‘y’. The training pair  $\{y^p, C_i^p\}_{p=1}^P$  is generated by dividing into ‘P’ patches, the coarse image and its corresponding HR classified map.

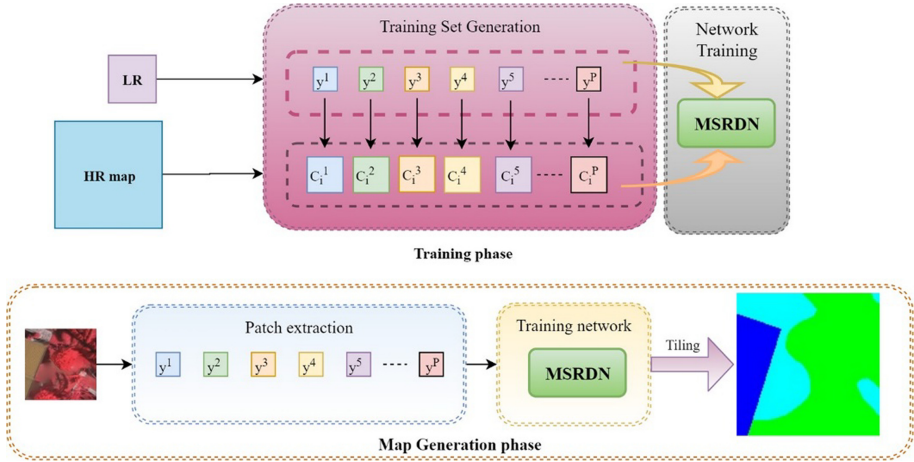


Fig. 1. Overview of Multi-Scale Residual Dense Network model for SRM application

### 2.2 MSRDN Network Training

The Fig. 2 shows the architecture of Multi Scale Residual Dense Network (MSRDN) comprising of Residual Dense Blocks (RDB) [14] at multiple scales as the name suggests, which extracts features at multiple level. The network is composed of three steps: **Shallow feature extraction** ( $F_{SF}$ ) from two successive convolution blocks containing Convolution, Batch Normalization and Relu layers, **Dense feature extraction** ( $F_{Di}$ ) at multiple scales by making use of the RDB Blocks and the Transition Upscaling (TU) Blocks.  $F_{SF}$  is fed as the input to both the stages and the presence of multiple stages can be justified by the order in which the TU Block and RDB blocks are brought to the vicinity of the shallow feature input. Either upscaling is followed by feeding of the corresponding output to the RDB Block or vice-versa. Finally, the **fusion of multi-scale dense features** ( $F_{MSD}$ ) by concatenation followed by dropout layer.

Let  $y_m$  be the input image patch for the network.

$$F_{SF} = H_{SFE}(y_m) \tag{1}$$

where  $H_{SFE}$  represents shallow feature extraction.  $F_{SF}$ , is the input to both the stages of the MSRDN net. First stage comprises of the entry of  $F_{SF}$  into the RDB Block ( $H_{RDB}$ ) after which upscaling is performed by TU Block ( $H_{TU}$ ). The swapping of these events

occurs at the second stage. The TU block consists of convolution layer with stride ‘2’ to upsample the input feature by a factor of ‘2’.

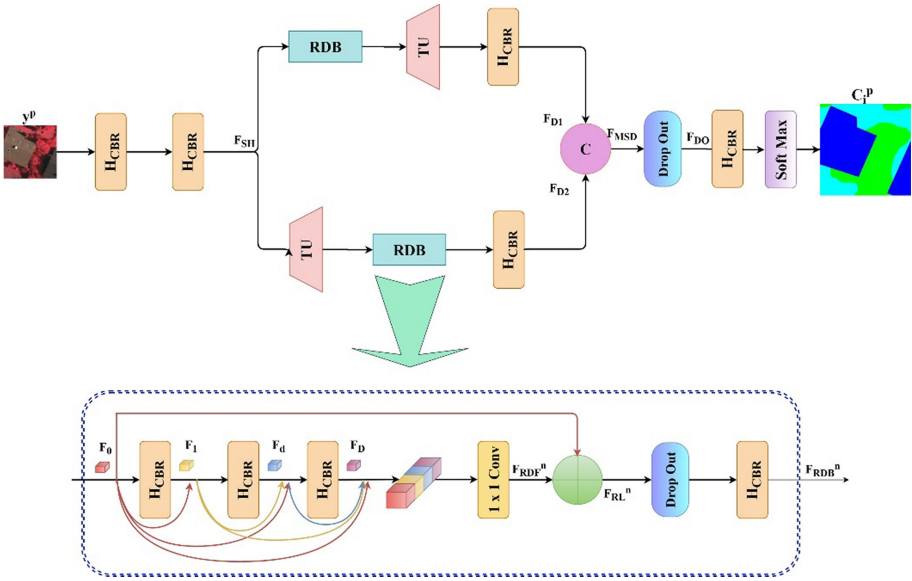


Fig. 2. Multi-Scale Residual Dense Network architecture for SRM application

RDB block comprises of residual connection of dense block, dense feature fusion ( $F_{RDF}$ ) and residual learning. The input to RDB block is denoted as  $F_0$  and the functionality of  $n^{th}$  stage RDB is represented as follows

$$F_{RDB}^n = H_{RDB}(F_0) \tag{2}$$

The dense block constitutes of dense connection of  $d$  convolution blocks.

$$F_{RDF}^n = H_{RDF}([F_0, F_1, F_d, \dots, F_D]); (1 \leq d \leq D) \tag{3}$$

where  $[..]$  represents concatenation of the features,  $H_{RDF}$  represents the functionality of fusion of dense feature within RDB block. The features from  $d$  convolution blocks are concatenated and are subjected to a  $1 \times 1$  convolution layer.  $F_d$  is the output of  $d^{th}$  convolution block of RDB block. Each convolution layer consists of a convolution layer followed by Batch normalization and RELU layer and its functionality is denoted by  $H_{CBR}$ .

$$F_d = H_{CBR}([F_0, F_1, \dots, F_{d-1}]) \tag{4}$$

The residual learning feature ( $F_{RL}$ ) is the sum of the input feature  $F_0$  and residual dense feature  $F_{RDF}$ . This is followed by dropout layer ( $H_{DO}$ ) and a convolution block ( $H_{CBR}$ ).

$$F_{RL}^n = F_{RDF}^n + F_0 \tag{5}$$

$$F_{RDB}^n = H_{CBR}(H_{DO}(F_{RL}^n)) \quad (6)$$

The outputs of both the stages are marked as  $F_{D1}$  and  $F_{D2}$  respectively.  $F_{D1}$  is the upscaled residual dense feature, whereas  $F_{D2}$  is the residual dense feature of the upscaled shallow feature.

$$F_{D1} = H_{CBR}(H_{TU}(H_{RDB}(F_{SH}))) \quad (7)$$

$$F_{D2} = H_{CBR}(H_{RDB}(H_{TU}(F_{SH}))) \quad (8)$$

After which fusion of  $F_{D1}$  and  $F_{D2}$  by concatenation is performed and  $F_{MSD}$  is obtained. The dropout layer receives  $F_{MSD}$ , which aids in removal of the correlated features and prevents overfitting of the network. Finally

$$F_{MSD}=[F_{D1}, F_{D2}] \quad (9)$$

$$C_m = H_{CBR}(H_{DO}(F_{MSD})) \quad (10)$$

### 3 Experimental Results and Discussion

This section analyzes the performance of the proposed network. It includes dataset description, the experimental setup used for training the network and finally the evaluation of the network and discussion on the results.

#### 3.1 Dataset Description

This work uses the standard dataset provided by ISPRS covered over Vaihingen, Germany. The data set is associated with 33 patches of different dimensions with an average dimension of  $2064 \times 2494$  and each of them corresponds to True OrthoPhoto (TOP), extracted from the larger TOP mosaic. The TOP and its ground truth are 9 cm which corresponds to both the ground sampling distance. TOP contains three bands *viz.*, Near infrared, red and green bands. Dataset have defined six categories *viz.*, Impervious surfaces, building, low vegetation, tree, car and clutter.

#### 3.2 Experimental Setup

Stochastic gradient descent (SGD) algorithm is used to minimize the objective function. The SGD weight decay is set to  $10^{-7}$ , momentum to 0.9 and mini-batch size to 16. The network training is done for 100 epochs and 544 iterations per epoch. The initial learning rate was set to 0.1 and decreased by a factor 10 at every 20 epochs and trained with NVIDIA GeForce GTX 1050 Ti GPU.

### 3.3 Network Evaluation

A total of 33 data patches is split into 60/40% for training and testing with 20 and 13 patches respectively. The training set is limited for training deep network hence the data patches are divided into overlapping patches with a stride of 70 and a uniform patch dimension of  $128 \times 128$  resulting in 8754 and 5767 disjoint training and testing patches. The TOP data is further downsampled at  $fds = 2$ . The network parameters *viz.*, the number of convolution layers (D) in the RDB dense block and the growth rate (G) is set to  $D = 3$  and  $G = 16$ .

The network evaluation is done by comparing it with Unet [15] and Segnet [16] based segmentation model. Since conventional SRM models use soft classified result as an input for mapping, this work finds it meaningful in using a modified semantic segmentation network for comparison. In this regard image segmentation models used for Remote sensing images are modified by including an extra decoding layer at the end to increase the resolution of segmentation result thus resulting in a SRmap. The comparative results of super resolution mapping networks are shown in Fig. 3. It is obvious from the observed result that proposed MSRDN has performed better than the Unet and Segnet.

### 3.4 Quality Metrics

The assessment is performed using the following quality metrics: Accuracy, IoU accuracy, and weighted IoU. Accuracy is the ratio of correctly classified pixels to the total number of pixels in that class, according to the ground truth.

$$Accuracy\ Score = \frac{TP}{TP + FN} \quad (11)$$

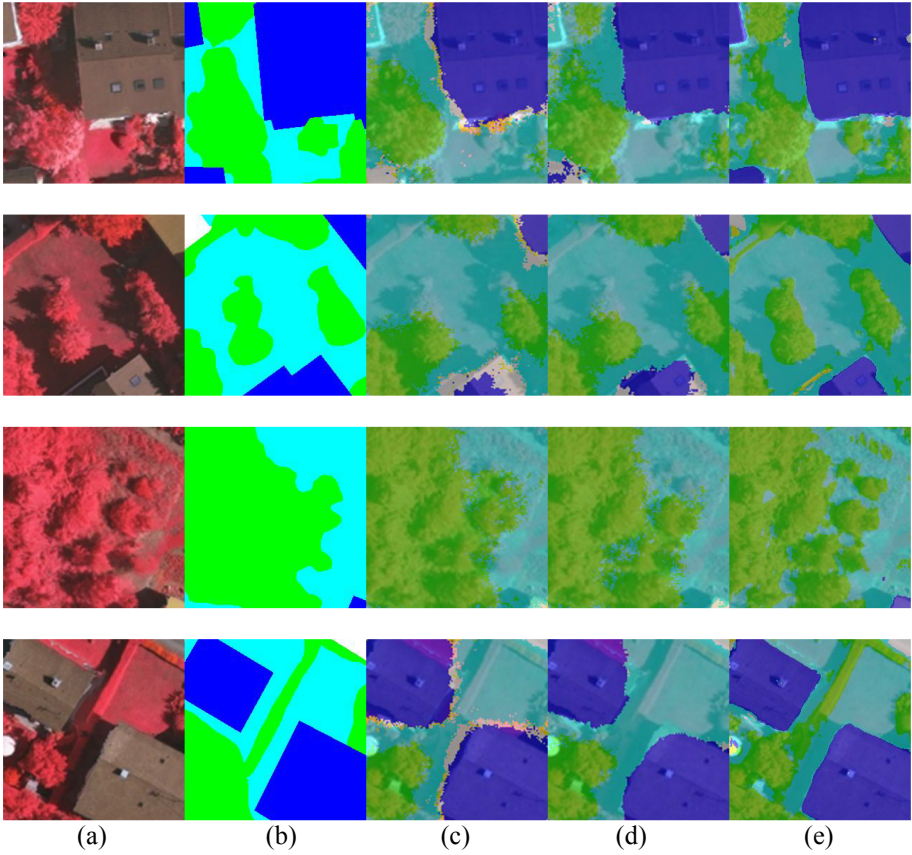
For the aggregate data set, Mean Accuracy is the average Accuracy of all classes in all images. Intersection over Union (IoU) accuracy is the ratio of correctly classified pixels to the number of ground truth and predicted pixels in that class *i.e.*,

$$IoU\ Score = \frac{TP}{TP + FP + FN} \quad (12)$$

where TP, FP and FN are the number of True Positives, False Positives and False Negatives. Weighted-IoU is the Average IoU of each class, weighted by the number of pixels in that class. This metric is useful when images have disproportionally sized classes, to reduce the impact of errors in the small classes on the aggregate quality score (Table 1).

### 3.5 Robustness to Ambiguities and Mislabeled Ground Truth

In the ISPRS dataset the provided ground truth suffers from few uncertainties resulting in erroneous labeled data. The ambiguities arise due to missing some objects while labelling, sharp transitions in labelled data for unsharp true data. This has led the trained network to overfit and misclassify compared with the ground truth but actually the network has performed better than the ground truth. This is obvious in Fig. 4. To explain



**Fig. 3.** Comparative results of super resolution mapping networks. Patches from (a) TOP, (b) HR ground truth, (c) Unet results overlaid, (d) Segnet results overlaid, and (e) MSRDN (proposed) results overlaid.

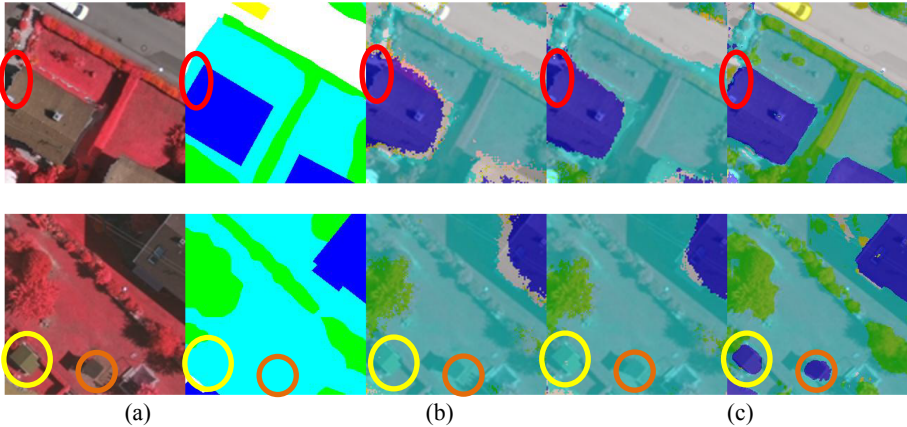
**Table 1.** Results on ISPRS Vaihingen dataset.

Methods	Global accuracy	Mean accuracy	Mean IoU	Weighted IoU	Training time (sec)
Unet	0.7389	0.6568	0.4021	0.4985	31149
Segnet	0.7564	0.6729	0.4356	0.5687	26517
MSRDN (proposed)	0.7732	0.7181	0.5281	0.6462	21093

better, there is visible presence of an impervious surface at the western corner of the true image (TOP) in Fig. 4(a). Ironically, the Ground Truth displays it as a part of low vegetation whereas the Unet and Segnet results map it as a building rather than labeling



it as an impervious surface. In the second case of ambiguity, two small buildings at the southern end in the True Image have been missed out by Ground Truth while Unet and Segnet has mapped it along with the low vegetation cover but MSRDN has correctly classified as building. The labeling in such scenarios is perfect pertaining to the results of the MSRDN, dethroning the errors committed by Unet, Segnet and Ground Truth.



**Fig. 4.** Ambiguous results. Patches from (a) TOP, (b) HR ground truth, (c) Unet results (d) Segnet results, and (e) MSRDN (proposed) results

## 4 Conclusion

Super Resolution Mapping (SRM) is a land cover mapping method that generates land surface cover maps at fine spatial resolution from a coarse spatial resolution Remote Sensing (RS) image. The performance of conventional SRM techniques is limited by the accuracy of the soft classifier result, thus this work proposes a novel method fully utilizing the power of deep learning to learn a non-linear relationship that exist between the coarse resolution image and high-resolution map. In this regard, a deep Multi-Scale Residual Dense Network (MSRDN) is proposed for SRM application of satellite data which extracts hierarchical features that can efficiently map sub-pixels to an accurate class. This network is trained with coarse resolution image and a fine resolution map patches to learn a super resolution mapping of land cover. The results of the experiments show that the proposed network provides an alternate solution to conventional SRM techniques, in generating a precise land cover information directly from low resolution images.

## References

1. Li, L., Chen, Y., Xu, T., Liu, R., Shi, K., Huang, C.: Super-resolution mapping of wetland inundation from remote sensing imagery based on integration of back-propagation neural network and genetic algorithm. *Remote Sens. Environ.* **164**, 142–154 (2015)

2. Li, L., et al.: Enhanced super-resolution mapping of urban floods based on the fusion of support vector machine and general regression neural network. *IEEE Geosci. Remote Sens. Lett.* (2019)
3. Thornton, M.W., Atkinson, P.M., Holland, D.A.: Sub-pixel mapping of rural land cover objects from fine spatial resolution satellite sensor imagery using super-resolution pixel-swapping. *Int. J. Remote Sens.* **27**(3), 473–491 (2006)
4. Ling, F., Du, Y., Li, X.D., Li, W.B., Xiao, F., Zhang, Y.H.: Interpolation-based super-resolution land cover mapping. *Remote Sens. Lett.* **4**(7), 629–638 (2013)
5. Mertens, K.C., De Baets, B., Verbeke, L.P.C., De Wulf, R.R.: A sub-pixel mapping algorithm based on sub-pixel/pixel spatial attraction models. *Int. J. Remote Sens.* **27**(15), 3293–3310 (2006)
6. Ling, F., Du, Y., Li, X.D., Zhang, Y.H., Xiao, F., Fang, S.M., Li, W.B.: Superresolution land cover mapping with multiscale information by fusing local smoothness prior and downscaled coarse fractions. *IEEE Trans. Geosci. Remote Sens.* **52**(9), 5677–5692 (2014)
7. Chen, Y.H., Ge, Y., Chen, Y., Jin, Y., An, R.: Subpixel land cover mapping using multiscale spatial dependence. *IEEE Trans. Geosci. Remote Sens.* **56**(9), 5097–5106 (2018)
8. Ling, F., et al.: Learning-based superresolution land cover mapping. *IEEE Trans. Geosci. Remote Sens.* **54**(7), 3794–3810 (2016)
9. Muad, A.M., Foody, G.M.: Impact of land cover patch size on the accuracy of patch area representation in hnn-based super resolution mapping. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **5**(5), 1418–1427 (2012)
10. Zhang, L.P., Wu, K., Zhong, Y.F., Li, P.X.: A new sub-pixel mapping algorithm based on A BP neural network with an observation model. *Neurocomputing* **71**(10–12), 2046–2054 (2008)
11. Zhang, Y.H., Du, Y., Ling, F., Fang, S.M., Li, X.D.: Example-based super-resolution land cover mapping using support vector regression. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **7**(4), 1271–1283 (2014)
12. Yang, X., Xie, Z., Ling, F., Li, X., Zhang, Y., Zhong, M.: Spatio-temporal super-resolution land cover mapping based on fuzzy C-means clustering. *Remote Sens.* **10**(8), 1212 (2018)
13. Ling, F., Foody, G.M.: Super-resolution land cover mapping by deep learning. *Remote Sens. Lett.* **10**(6), 598–606 (2019)
14. Zhang, Y., Tian, Y., Kong, Y., Zhong, B., Fu, Y.: Residual dense network for image super-resolution. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2472–2481 (2018)
15. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) *MICCAI 2015*. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28)
16. Audebert, N., Le Saux, B., Lefèvre, S.: Semantic segmentation of earth observation data using multimodal and multi-scale deep networks. In: Lai, S.-H., Lepetit, V., Nishino, K., Sato, Y. (eds.) *ACCV 2016*. LNCS, vol. 10111, pp. 180–196. Springer, Cham (2017). [https://doi.org/10.1007/978-3-319-54181-5\\_12](https://doi.org/10.1007/978-3-319-54181-5_12)