Generating annual high resolution land cover products for 28 metropolises in China based on a deep super-resolution mapping network using Landsat imagery

Da He\textsuperscript{ab}, Qian Shi\textsuperscript{ab}, Xiaoping Liu\textsuperscript{ab}, Yanfei Zhong\textsuperscript{c}, Guisong Xia\textsuperscript{c} and Liangpei Zhang\textsuperscript{c}

\textsuperscript{a}School of Geography and Planning, Sun Yat-sen University, Guangzhou, China; \textsuperscript{b}Guangdong Provincial Key Laboratory for Urbanization and Geo-simulation, Guangzhou, China; \textsuperscript{c}State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China

**ABSTRACT**

High resolution of global land cover dynamic is indicative for understanding the influence of anthropogenic activity on environmental change. However, most of the land cover products are based on Landsat image that only has 30 m resolution, which is insufficient to distinguish the heterogeneous urban structure; while very high spatial resolution image usually has low temporal resolution, which is difficult to monitor the urban dynamic. Deep-learning-driven super-resolution mapping is a prevailing way of achieving very-high-resolution land cover dynamic products in aspect of alleviating the mixed pixel problem of Landsat image. However, two limitations are obvious: 1) the fixed grid of kernel during the upsampling process favors spatial homogeneity and suppresses the learning of spatial heterogeneity of urban composition and 2) geometric or radiation variation over large spatial and long temporal extent in remote sensing images makes the super-resolution mapping approach difficult to transfer for application. Here, we attempt to solve these two limitations: 1) a progressive edge-guided super-resolution architecture is designed to allow nonuniformed kernel specific at the low-confidence edge region and intensify the learning of heterogenous compositions’ patterns and 2) an alternating optimization strategy is designed to minimize the resultant entropy and modulate the classification hyperplane to accommodate to the manifold of the discrepant region. Validation experiments are investigated based on a fine-grained and large-extent super-resolution (FLAS) dataset constructed in this study, and it is found that our approach remarkably enhances rich detailed patterns of heterogenous region and outperforms other state-of-the-art algorithms. Besides, we applied DETNet to the large spatial extent of 28 metropolises in China (>40,000 km\textsuperscript{2}) and the large temporal extent of continuous 21-year (2000–2020) in Wuhan city to examine transferability. From the land cover areas variation, we find that the expansion rate of cropland is faster than the urban expansion over the past 10 years, which are gradually becoming the principal source for the encroachment of forest and lakes. From detailed urban dynamic reflected by the 21-year products, we find that urban-villages between the old city zone and the outer high-tech development zone are gradually disappeared. The captured dynamic is consistence with the urban-village renovation policy during this period, which is meant to redistribute the spatial configuration of the city for a more sustainable urban structure. We believe that the proposed method can facilitate a seamless and fine-grained observation system that can fill the weakness of the existing land cover activities and provide a brand-new insight into the urban dynamic and its underlying mechanism.

1. Introduction

Land cover information retrieval through remote sensing imagery is a promising way for understanding the socioeconomic properties and interactions between human activities and environmental change (Liu et al. 2019, 2020). Benefit to its long time-series and global coverage, remote sensing imagery can provide an unprecedented insight into the progress of sustainable development (Burke et al. 2021), afforestation (Wernick et al. 2021), and urbanization (Zhang, Chen, and Lu 2015; Deng and Zhu 2020; Aljaddani, Song, and Zhu 2022; Zhu et al. 2022) and help municipality make informed decisions on urban green space management (Chen et al. 2021a), coastal ecosystem monitoring (Li et al. 2019), water/wetland resource management (Pekel et al. 2016; Hu et al. 2021), and environment protection (Aguirre-Gutiérrez et al. 2020).

Land cover information retrieval usually undertaken by land cover classification and has been exhausted by pioneer, achieving many land cover products, such as
the 1-km Global Land Cover Characterization (GLCC) product in 1992 (Loveland et al. 2000), the 1-km Global Land Cover (GLC2000) product in 2000 (Bartholome and Belward 2005), the 1-km Global Land Cover dataset provided by National Mapping Organizations (GLCNMO) product in 2003 (Tateishi et al. 2011), the 300-m Global Land Cover product (GlobCover) in 2005, 2006, and 2009 (Bicheron et al. 2011), the 500-m annual MODIS Global Land Cover type (MCD12Q1) product (Friedl et al. 2002), the 30-m Finer Resolution Observation and Monitoring-Global Land Cover (FROMGLC30) product in 2010 and 2015 (Gong et al. 2013), the 30-m GlobeLand30 product in 2000, 2010, and 2020 (Chen, Ban, and Li 2014), the 30-m National Land Cover Database (NLCD) product at USA every 5 years since 2006 (Fry et al. 2011), the 30-m GLC_FCS30 product in 2020 (Zhang et al. 2021a), and the 10-m Esri Land Cover product in 2020 produced by European Space Agency (ESA) (Karra et al. 2021).

Although the spatial resolution is constantly increased, the existing land cover products are still not enough to delineating the spatial heterogeneity of the highly dynamic urban compositions (Deng and Zhu 2020). For example, the diameter of the urban tree crown is usually less than 10 m, the width of the non-primary road is usually shorter than 30 m. Even the length of the small residential buildings is larger than 30 m, they may also be merged with the surroundings, leading to the destruction of their morphological structure. In this situation, we find that one Landsat pixel (spatial resolution of 30 m) may capture a mixture of the reflectance signal of multiple land cover types, resulting in the mixed pixel problem (Fisher 1997). Conventional pixel-level hard classification approach adopted by existing products usually simplified each pixel grid as one single land cover class (Adams et al. 1995), which ignored the spatial structure of the mixed land covers within 30 × 30 m ground area, leading to a large amount of latent land cover information loss and fine-grained structure loss (Wang et al. 2020).

In the past decades, the mixed pixel problem of Landsat image in urban scene has been thoroughly studied using spectral unmixing (SU) technique, which establishes physical mixture model to simulate the reflected signal mixing process, and try to decompose it to several pure signals (called endmember) and inversing their proportions (called abundance) using least square optimization. The output of SU is a soft classification map (called abundance map), i.e. each pixel is proportional belonging to each endmember class, instead of hard classification that assigning one mixed pixel to a single class. SU has been widely used to explore the proportion of impervious surface or forest inside each mixed pixel (Xiao and Moody 2005; Deng et al. 2017). However, the abundance maps are still retained at 30-m resolution, which only provide fraction indications without any spatial distribution of each land cover class inside a mixed pixel.

In order to achieve high-resolution (HR) land cover dynamic products, in this study, we attempt to use the super-resolution mapping (SRM, also known as sub-pixel mapping) approach (Atkinson 1997) for low-resolution (LR) Landsat image. Specifically, SRM split each mixed pixel (30 × 30 m ground area) into sub-pixel locations (circa 4 × 4 m ground area in our study) and assign the land cover classes on each sub-pixel location to restore a sub-pixel-scale land cover distribution and realize spatial resolution enhancement (Atkinson 1997). The assignment principal follows the proportion indication derived from SU (Keshava and Mustard 2002; Bioucas-Dias et al. 2012), and the location indication derived from the prior knowledge of the spatial distribution (Wang et al. 2021b). SRM has been widely applied in practical application that requires meticulous measures. For example, Ardila et al. (2011) applied SRM approach on 2.4-m QuickBird MS image at Bothoven, Enschede, to reconstruct a 0.6-m urban tree cover maps with individual plants distinguished; Muad and Foody (2012) used SRM approach to super-resolution the 240-m remote sensing image at Quebec province, Canada to 30 m for mapping the lake area; Poudyal (2013) used SRM approach to super-resolve the 2-m WorldView-2 image to 0.5-m resolution for mapping the individual plants in the potato farm field and monitoring their growing status; Ling et al. (2019a) applied SRM on 463-m MODIS image at Bay of Bengal to reconstruct a 30-m resolution map to measure the narrow river wetted width; Zhang et al. (2019) applied SRM approach on 250-m MODIS images at forest areas in the USA, Paraguay and Russia, to reconstruct time-series 25-m resolution forest maps, and discovered tiny changes that was caused by the forest fire and deforestation; He et al. (2022) developed a novel SRM approach on 10-m Sentinel-2 images to derive time-series 2-m urban tree covers that are usually omitted in traditional land cover/forest cover product, etc.
Generally, SRM relies on spatial prior knowledge to regularize the ill-posed problem of reconstructing LR to HR image (Atkinson 2009), which provides the indications of how each class located within a mixed pixel. We summarized four types of spatial prior: 1) Spatial dependence prior assumes that spatial proximate pixels more likely belong to the same class than the distant one, which includes spatial attraction model (Mertens et al. 2006), pixel swapping model (Atkinson 2005), genetic algorithm SRM (He et al. 2016a), Hopfield neural network SRM (Su 2019), spectral-spatial fusion-based SRM (Xu et al. 2018), and random field-based SRM (Kasetkasem, Arora, and Varshney 2005); 2) Geostatistical prior assumes that the sub-pixel distribution is subject to empirical geostatistical models that can be regressed with given samples (Boucher and Kyriakidis 2006), which includes multi-point simulation (Ge 2013), area-to-point kriging (Wang et al. 2015), and spatial distribution pattern-based SRM (Ge et al. 2016); 3) Regularization prior assumed that HR image follows the specific manifold like Laplacian model, total variation model, non-local prior or sparse prior, which includes maximum a posteriori SRM (Zhong et al. 2015), and sparse representation SRM (Song et al. 2019); 4) Spatiotemporal prior is aimed to provide complementary information from external dataset for constraint, such as time-series temporal images (Li et al. 2017), shifted images (Zhong et al. 2014), panchromatic images (Nguyen, Atkinson, and Lewis 2011), and Point of Interest (POI) (Chen et al. 2018b). However, the spatial priors aforementioned are usually model-driven or expertly handcrafted, which is too inflexible for reconstructing the irregular heterogeneous distribution (Ling et al. 2019b). Recently, deep-learning-based super-resolution mapping network (DLSMNet) is developed, which is aimed to learn a more realistic sub-pixel-scale spatial priors from exemplar pairs of LR image and HR annotations based on the end-to-end mapping and representation learning ability of the network (Yu et al. 2017). Many prevailing network architectures were taken as backbone to formulate advanced variants of DLSMNet, such as super-resolution reconstruction network (Ling et al. 2019a; Ling and Foody 2019b; Ma et al. 2019; He et al. 2021a; Shang et al. 2020), semantic segmentation network (Arun, Buddhiraju, and Porwal 2018; Jia et al. 2019), graph convolution network (Zhang et al. 2021b), and spatiotemporal fusion network (Chen, Shi, and Ge 2021b). The efficient network modules such as attention mechanism (He et al. 2021b, 2022) and multi-scale pyramid (Wang et al. 2019) were also fully explored, and the interpretable network modeling was also studied by embedding physical model of SU and SRM into the network (He et al. 2021c; Yin et al. 2021).

Although the DLSMNet is a prevailing trend in SRM community, it leaves several open questions to be further studied: 1) The network used stacked convolutional layers to enlarge the receptive field for learning high-level feature, which is equivalent to downsampling and sacrifices the low-level spatial details (Huan et al. 2022). Traditional DLSMNet usually adopted the fixed grid sampling of $3 \times 3$ (or other size) kernel in the deconvolution or sub-pixel convolution layer to restore the downsampled feature map; however, fixed grid sampling favors spatial homogeneity and makes the spatial details merged with the neighboring dominant class and being ruined (Marin et al. 2019); 2) Landsat images in different regions may have geometrics or radiance variations due to the different acquisition conditions (Tuia, Persello, and Bruzzone 2016), which result in a feature distribution gap between training data (source domain) and test data (target domain) (Kouw and Loog 2018). The network only trained on the training set may confront overfitting problem and suffer a performance drop on the large-extent Landsat images. Although a large amount of training data can strengthen the generalization performance (Kolesnikov et al. 2020), the meter-level high resolution annotation data are insufficient and not adequate in quality (Tong et al. 2020). Therefore, generalizing the network to the diverse large extent images remains a major challenge.

In this study, we developed a novel deep edge-guided and transferable super-resolution mapping network (DETRNet) to alleviate the weakness of the fixed grid sampling in reconstructing spatial heterogeneity and improve the generalization ability of the network model, achieving HR land cover products over large-extent and over time. The main contributions are as follows: 1) we designed an architecture in DETNet to adaptively search the error-prone pixel for locating the potential edge region using $1 \times 1$ multiple layer perceptual (MLP), which enables a nonuniformed sampling of convolution that adhered to the edge to intensify the learning of the edge region; 2) we designed a strategy to alternatingly minimize the target
domain’s entropy and the source domain’s loss function during model fine-tuning, to gradually project the classification hyperplane to the manifold of the target domain, making the DETNet model valid on Landsat image in large-extent and different time; 3) in order to train a generalized DETNet, a fine-grained and large-extent sub-pixel-scale annotation (FLAS) dataset is constructed by surveying and mapping experts, with pairs of 30-m Landsat image and very-HR annotation, which have a total coverage of more than 6362 km². After that, we used the trained DETNet to reconstruct 30-m Landsat image to meter-level HR land cover maps with rich geographical-realistic details, over large extent of the 28 metropolises (>40,000 km²) in China, and over long time series of continuous 21 years (2000–2020) in Wuhan city, to examine its feasibility in understanding how cities change in space and over time.

2. Dataset

2.1. Training dataset and pre-processing

Up to now, meter-level well-annotated land cover dataset is significantly deficient. Although several HR annotated datasets exist in the community (Gerke et al. 2014; Maggiori et al. 2017), the coverage of which are limited within 10 km² and the geographic patterns are similar, which is insufficient for training a generalized model suitable for large extent mapping. Although the Gaofen-2 Image Dataset (GID) annotation has both high spatial resolution and the large spatial extent, i.e. 4 -m resolution and 50,000 km², the annotation is not sufficiently fine-grained, as shown in Figure 1. To this end, this research firstly constructed a Fine-grained and Large-extent SRM HR annotation (FLAS) dataset.

Specifically, 211 annotation locations are randomly sampled in Guangdong province, China, and the Google Earth images with size of nominal 4500 × 6700 pixels at 1-m resolution acquired in these sampling locations in summer 2019 are used (Figure 1). The sampling region covers more than 6362 km², which enables a diversity of spatial distribution patterns to ensure the generalization of the trained model. The pixel-wise annotation work was carefully done on the HR Google Earth images with the help of experts major in survey and mapping. Since they usually take ground survey within Guangdong province and are familiar with the in-situ ground truth, they have prior knowledge of which parcel on the image should be annotated as forest, shrub, grass, etc. As for annotation implementation, we manually delineate the boundary vector of each land cover parcel on the HR Google Earth images with the aid of ArcGIS. For the land cover parcel that is hard to determine, the NDVI, NDWI, NDBI, and DEM are also taken as the assistant data for a reliable determination. After that, the delineations are double checked by another group without participated in the previous work, to further screen and correct the poor-quality annotation. Finally, the boundary vectors are converted to raster

![Figure 1](image-url) The sampling location of the FLAS annotation dataset. One scene of the FLAS annotation is enlarged for visual display, and the corresponding GID annotation at the same field of view is also displayed for comparison.
image to formulate the FLAS annotation label. Considering the balance between super-resolution scale and reliability, and in order to stay in step with the GID annotation dataset, the FLAS annotation dataset is resampled to 4-m resolution to release the reconstruction difficulty and ensure the reliability. The seven discriminant categories are chosen based on the Chinese Land Use Classification Criteria (GB/T 210102017), i.e. cropland, forest, grass, shrub, water, impervious surface area (abbreviated as impervious in the following), and bare land. In addition, this annotated dataset was examined by third party that is not involved in annotation work, for a further refinement. The number of the annotated pixels and the ratio for each category is provided in Figure 2.

The training of the SRM network requires pairs of LR image and HR annotation dataset, thus the Landsat-8 Operational Land Imager (OLI) image with 30-m resolution is selected. Specifically, Landsat-8 OLI top-of-atmosphere (TOA) images deposited in Google Earth Engine (GEE) (https://developers.google.com/earth-engine/datasets/catalog/LANDSAT_LC08_C02_T1_TOA#description) is used in our study to avoid atmospheric effect and ensure consistent observation in large area, as was done by Xiong et al. (2017). In order to eliminate the cloud contamination and the influence of phenology factor, the Landsat images within the growing period of April–September 2019 are composited by median-mosaic operation (https://developers.google.com/earth-engine/apidocs/ee-imagecollection-median), and totally seven bands are used. The Google Earth images and the corresponding images are projected to WGS-84 coordinate and were rigorously registered using ENVI, with registration error less than a Google Earth image pixel.

We split the 211 FLAS annotated scene images into training set, validation set, and test set based on the proportions of 70%, 10%, and 20%, i.e. 147, 22, and 42 scene images, respectively, which are denoted as red, blue, and green block in Figure 1 and non-overlapped with each other. The patches of size 512 × 512 on the FLAS annotations and the corresponding Landsat patches of size 64 × 64 on the Landsat images were cropped to form pairs of HR and LR samples, which can be fed to the network for training a SRM task model. Totally, the 1536 patches (512 × 512 pixels) are cropped from the 147 scenes images (circa 1125 × 1675 pixels), 233 patches are cropped from the 22 scene images, and 485 patches are cropped from the 42 scene images. The specific cropping procedure can be referred to the supplementary file. The training set is used to train the network model, the validation set is used to evaluate the model at each training epoch to select the optimal network model, and the test set is withheld until the final test, which provide the final accuracy of the network.

### 2.2. Study area

In order to examine the large-extent generalization ability, 28 provincial capital metropolises in China are selected as the study areas, which are located at the temperate monsoon climate and the subtropical monsoon climate zone, ranging from 19 to 46° N, 102 to 130° E, as shown in Figure 3. The residual 6 provincial capital
cities are located at the continental climate and mountain climate zone, which have large variation in land cover types and statistic distribution characteristic.

In consideration of the importance of these provincial capital cities, a regular observation and land cover mapping is necessitated for an in-time monitoring of the land cover changes and analyzing their internal driving force and influence to the eco-systems. It helps understand the interactions between human activities and environmental change, which is important for realizing the goal of sustainable development.

3. Method

3.1. Sub-pixel mapping (SRM)

SRM is a widely used technique for mitigating the mixed pixel problem by reconstructing the geographical locations of each land covers at sub-pixel-scale inside the mixed pixel. Classical SRM usually relies on the abundance images (soft classification, i.e. one pixel proportionally belongs to multiple class) derived from SU technique (Bioucas-Dias et al. 2012) and further specifies the sub-pixel-scale locations of each land cover. Given a soft classification of \( 3 \times 3 \) mixed pixels with three classes (Figure 4), SRM grid divides the mixed pixels into several sub-pixel sites by interpolation according to the scale factor \( S \) (in this case \( S=5 \), i.e. one mixed pixel is split into 25 sub-pixels) and allocates the land cover labels to each sub-pixel site, in accordance with the proportion indication. This realizes a hard classification (i.e. one sub-pixel belongs to only one class) at sub-pixel-scale with spatial resolution improved by \( S \). Compared to the traditional pixel-scale hard classification assigning one pixel to single label (Figure 4), more morphological structure and detailed edges can be reflected by SRM result.

Given the Landsat image \( Y = \{ y_i \in \mathbb{R}^{c \times 1}, i = 1 \ldots N \} \), \( (N = h \times w, h, w, c \) represents the height, width, and band of the image). \( y_i \) represents the \( i \)th mixed pixel covering \( 30 \times 30 \) m ground footprint, and its reflectance signal is assumed to be a combination of the signals of multiple pure land cover types. This mixed reflectance signal can be represented by the sum of the endmember’s spectral signal \( M \in \mathbb{R}^{c \times NC} \) weighted by their abundance \( a \in \mathbb{R}^{NC \times N} \) based on the SU technique (NC represent the number of land cover types), given as:

![Figure 3. Study area of the 28 metropolises in China.](image-url)
where $a_{i,k}$ represent the proportion of the $k$th class in the $i$th mixed pixel, which should follow the abundance nonnegative constraint (ANC) and abundance sum-to-one constraint (ASC) to restrain the proportion value to conform to the physical meaning, i.e. proportion value should be no less than 0 and total proportion values in one mixed pixel should equal to 1. However, SU only provides abundance information, without any indication of where each endmember spatially located in the mixed pixel.

SRM is then adopted to specify the spatial location of each endmember at sub-pixel scale based on its spatial dependence with the surrounding pixels (e.g. eight-neighborhood) and finally extend the LR abundance image $\alpha \in \mathbb{R}^{NC \times N}$ to the HR classification map $X \in \mathbb{R}^{1 \times (N \times S^2)}$. SRM is formulated based on the observation model to bridge the LR and HR image (Zhong et al. 2015). Considering that the metric criterion of $X \subset \{1, 2, 3, \ldots, NC\}$ and the $\alpha \subset [0, 1]$ is not compatible, one-hot operation is used to uniformed two by transforming the $X$ to the probability map $Z \in \mathbb{R}^{NC \times (N \times S^2)}$. In this way, the observation model of SRM problem can be given as:

$$a = ZD$$

(2)

where $D \in \mathbb{R}^{(N \times S^2) \times N}$ is the downsampling matrix, which represent the degradation process from $Z$ to $a$.

Substitute the Eq. (2) into Eq. (1) can formulate the super-resolution task of super-resolution LR Landsat image to HR classification map:

$$Y = M\alpha + n, s.t.\text{ANC}: a_{i,k} \geq 0, \text{ASC}: \sum_{k=1}^{NC} a_{i,k} = 1$$

(1)

$$Y = MZD + n, s.t.\text{ANC}: Z_{i,k} \geq 0, \text{ASC}: \sum_{k=1}^{NC} Z_{i,k} = 1$$

(3)

Considering that the traditional model-driven or expertly handcrafted priors is difficult to estimate $D$ due to the irregular spatial patterns, Ling et al. (2019b) argued to learn geographical-realistic priors by data-driven deep learning method and developed deep-learning-based super-resolution mapping network (DLSMNet). Given training pairs of $(Y, X)$, the $M$ and $D$ can be implicitly estimated as network parameter $\theta$, by minimizing the distance $\mathcal{L}$ between the network predicted result $\hat{X}$ and the ground truth $X$, give as:

$$\hat{X} = f_{network}(Y; \theta)$$

(4)

$$\hat{\theta} = \arg\min_{\theta} \{ \mathcal{L}(\hat{X} - X) \}$$

(5)

### 3.2. Deep edge-guided and transferable Sub-pixel Mapping Network (DETNet)

Considering the limitations of the current DLSMNet in spatial detail reconstruction and transferability, this research developed DETNet model. Generally, DETNet consists of three main blocks according to its purpose (Figure 5), i.e. feature representation block is for extracting discriminative feature for the mixed land covers, upsampling block is for mapping the features from LR to HR feature space with the proposed PESR architecture, and loss function block is to measure the difference between the estimated HR classification and the HR ground truth. At the training stage, the gradient
3.2.1 Feature representation with ResNet101

Firstly, given coarse input Landsat patch \( Y \in \mathbb{R}^{H \times W \times C} \), it is upsampled to the target size \( Y' \in \mathbb{R}^{H' \times W' \times C} \) (where \( H = h \cdot S, W = w \cdot S \)) to set up sub-pixel locations, by bilinear interpolation layer. After that, one of the most commonly used CNN architecture, Residual Network with 101 layers (ResNet101) is adopted as the feature representation block (He et al. 2016b), to extract discriminative features of the mixed land cover for further modeling. The specific structure of ResNet101 is provided in Table 1.

The convolutional transformation of ResNet101 on \( Y' \) is given as:

\[
\text{Conv}(Y') = \sigma(W \otimes Y' + b) 
\]

\[
\text{ResNet}(Y') = \text{Conv}(\ldots \text{Conv}(\text{Conv}(Y'))) 
\]

where \( W \) and \( b \) are weights and biases, which are the network parameters to be solved.

The output feature maps of ResNet101 (Layer 4, denoted as \( F_{4} \in \mathbb{R}^{H'/16 \times W'/16 \times 1024} \)) is taken as the high-level semantic features, which is fed to Atrous Spatial Pyramid Pooling (ASPP) layer and SoftMax classifier layer to derive the preliminary class probability map \( F_{mp} \in \mathbb{R}^{H/8 \times W/8 \times NC} \), as was done by Chen et al. (2018a).

\[
F_{mp} = \text{SoftMax}(\text{ASPP}(\text{ResNet}(Y'))) 
\]

3.2.2 Progressively Edge-guided Super-resolution Reconstruction (PESR)

In order to reconstruct the geographical-realistic detail of the land covers, the PESR architecture is designed in upsampling block (Figure 6). The underlying mechanism of PESR is to adaptively localize the fine-grained edges at multiple scale and refine the prediction of these detailed edges based on the additionally trained nonuniform point-wise classifier. Noteworthy is that the edge enhancement has been thoroughly studied in deep learning community; however, most of which are based on redundant dual branch structure and rely on extra prepared edge annotations for supervision (Yang et al. 2017). On the contrary, our method is free from extra edge
annotations, since it can adaptively search the error-prone pixel for locating the potential edge region.

Firstly, we design a progressive super-resolution architecture to mitigate the difficulty of filling a large scale gap at one time (Pan et al. 2020). Based on this architecture, the input \(F_{\text{inp}}\) is progressively upsampled by 2-fold once a time with sub-pixel convolution (SPC) layer (Shi et al. 2016), i.e. in the 5-fold upsampling stage \((S=2,4,8)\), the \(F_{\text{inp}}^{S/2}\) is upsampled to generate preliminary \(F_{\text{inp}}^{S}\). Meanwhile, in each upsampling stage, four steps are designed to refine the \(F_{\text{inp}}^{S}\) (Figure 7):

1 **Hard point sampling (HPS).** HPS strategy is to precisely localize the fine-grained edges of the land covers on \(F_{\text{inp}}\). Generally, the edges are the most ambiguous locations in a scene, with low confidence points that usually have indistinguishable probabilities value among each class (Dai et al. 2007). Based on this assumption, we measure the hard point by the difference between the largest and the second largest probability value point-by-point across the \(F_{\text{inp}}^{2}\). The measurement can be given as:

\[
\max \left( F_{\text{inp}}^{2}(i,j,:) \right) - \max_{\text{second}} \left( F_{\text{inp}}^{2}(i,j,:) \right) < \varepsilon \quad (9)
\]

where \(\max \left( F_{\text{inp}}^{2}(i,j,:) \right)\) means the largest probability value in the vector at point \((i,j)\), \(\max_{\text{second}} \left( F_{\text{inp}}^{2}(i,j,:) \right)\) means the second largest probability value. If the difference is less than threshold \(\varepsilon\), the point \((i,j)\) is selected. Totally, the \(N\) most ambiguous hard point are localized, forming a location set \(P\), where \(P_n \left[ F_{\text{inp}}^{2}(i,j,:) \right] (n = 1, \ldots, N)\) means the \(n\)th point of set \(P\) on \(F_{\text{inp}}^{2}\). A zoomed-in area is visualized in Figure 6 to display the localized hard points (red point).

2 **Spatial-semantic synergized features extraction.** The spatial-semantic synergized features of the fine-grained edges were extracted for the subsequent reclassification. Specifically, the “Layer 3” of ResNet101, denoted as \(F_{\beta} \in \mathbb{R}^{H/8 \times W/8 \times 512}\), is adopted as the complementary spatial features, which is

| Layer name | Output size | Kernel size |
|------------|-------------|-------------|
| Layer 1   | \(H/2 \times W/2 \times 64\) | 1 \(\times\) 1, 64, stride 2 |
| Layer 2   | \(H/4 \times W/4 \times 256\) | 3 \(\times\) 3, max pool, stride 2 |
| Layer 3   | \(H/8 \times W/8 \times 512\) | \(1 \times 1, 128\) |
| Layer 4   | \(H/16 \times W/16 \times 1024\) | 1 \(\times\) 1, 256 |

Figure 6. The detail architecture PESR.
merged with the semantic probability map \( F_{\text{map}} \) to generate spatial-semantic synergized features:

\[
P[F_2^2] = up_{\times 2}(\text{Concat} [P[F_1^3], P[F_{\text{map}}]])
\]  

where \( up_{\times 2}(\cdot) \) denotes upsampling by 2-times, \( \text{Concat}(\cdot) \) denotes concatenating two tensors along the channel dimension, and thus the channel dimension of \( P[F_2^2] \) is extended to \((512 + NC)\).

### 3. Nonuniform point-wise refinement (NPR).

The NPR is designed to reclassify the fine-grained edges by the additionally trained classifier, based on the extracted spatial-semantic synergized features \( P[F_2^2] \). Notably, instead of \(3 \times 3\) convolution, NPR classify the fine-grained edge points with a \( \text{MLP} \) classifier point-by-point, which can avoid the interference of the neighboring pixel introduced from the fixed grid sampling of the convolutional kernel (Kirillov et al. 2020). Specifically, \( \text{MLP} \) classifies each point \( P_n[F_c^2] \in \mathbb{R}^{1 \times 1 \times (512 + NC)} \) by transforming them into a semantic probability vector \( P_n[F_{\text{map}}^2] \in \mathbb{R}^{1 \times 1 \times NC} \):

\[
P_n[F_{\text{map}}^2] = \text{MLP}(P_n[F_c^2]), \; n = 1, \ldots, N
\]  

### 4. Replacing the point with the refined prediction.

The \( F_{\text{map}}^2 \) is refined by replacing the probability vector of the corresponding points on \( F_{\text{map}}^2 \) with the newly predicted \( P_n[F_{\text{map}}^2] \):

\[
F_{\text{map}}^2 = \text{NPR}(P_n[F_{\text{map}}^2], P[up_{\times 2}(F_{\text{map}})])
\]  

where \( \text{NPR}(\cdot) \) means replace the edge's probability vector on \( up_{\times 2}(F_{\text{map}}) \) with the reclassified one. The residual upsampling stages are the same way.

Finally, the outputs of \( \text{PESR} \) are the refined \( F_{\text{map}}^8 \) and \( F_c^8 \), which are denoted as \( \hat{Z} \in \mathbb{R}^{H \times W \times NC} \) and \( F_{\text{out}} \in \mathbb{R}^{H \times W \times (512 + NC)} \) for subsequent implementation, respectively. \( \hat{Z} \) represents the super-resolved HR probability maps, which can be hardened to obtain the sub-pixel-scale classification result \( \hat{X} \).

#### 3.2.3 Loss function

For training the network model, the estimated probability \( \hat{Z} \) is rectified by the ground truth \( Z \in \mathbb{R}^{H \times W \times NC} \) with cross entropy loss function, to supervise the learning of sub-pixel-scale classification:

\[
\text{SRM Loss} = \sum_{i=1}^{N} \sum_{j=1}^{W} \sum_{k=1}^{NC} Z(i, j, k) \cdot \log(\hat{Z}(i, j, k))
\]  

where \( \hat{Z}(i, j, k) \) denotes the predicted probability value at \((i, j)\) location in \( k \) channel and \( Z(i, j, k) \) denotes the real probability value at \((i, j)\) location in \( k \) channel.

Besides, in order to train the \( \text{MLP} \) classifier, the feature vectors at hard points \( P[F_{\text{out}}] \) are preliminarily classified to generate probability vectors, which are then rectified by the corresponding point in ground truth \( Z \) (denoted as \( P[Z] \)) with cross entropy loss function, given as:

\[
\text{Point Loss} = \sum_{n=1}^{N} \sum_{k=1}^{NC} P_{n,k}[Z] \cdot \log(\text{MLP}(P_{n,k}[F_{\text{out}}]))
\]
where \( P_{n,k}[Z] \) denotes the real probability value of the \( nth \) hard point in \( k \) channel and \( MLP(P_{n,k}[F_{out}]) \) denotes the estimated probability values of the \( nth \) hard point in \( k \) channel. Notably, the weights are shared across all MLP, as was done by Kipf and Welling (2017), which is simple and computational efficiency. Thus, the total loss function can be obtained by the weighted sum of \( SRM\_loss \) and \( Point\_loss \), given as:

\[
L_{seg} = Point\_Loss + \lambda \cdot SRM\_Loss
\]

(15)

Given \( M \) pairs of training sample \( \{(Y_m, X_m), m = 1 \ldots M\} \), \( X_m \in \mathbb{R}^{H \times W} \) is firstly one-hot encoded to generate ground truth probability map \( Z_m \in \mathbb{R}^{H \times W \times NC} \), and the network parameters \( \theta \) can be obtained by iteratively minimize the difference between the estimated \( \hat{Z}_m \) and ground truth \( Z_m \), with loss function \( L_{seg} \) as criteria, given as:

\[
\hat{\theta} = \text{argmin}\{L_{seg}(\hat{Z}_m, Z_m)\}, m = 1 \ldots M
\]

(16)

After converging, the trained model with parameter \( \hat{\theta} \) can be used to reconstruct the external Landsat images to generate sub-pixel-scale HR classification image.

### 3.2.4 Alternating optimization with entropy-minimization strategy (AOE)

Although well-trained on the training sample (source domain image), the network model may also be confronted with a performance drop on the Landsat image at unknown areas (target domain image) due to the feature distribution gap between source and target domain, which impedes the usability of DETNet in large extent mapping. In order to reduce the cross-domain discrepancy between source and target results, a criterion that measures this discrepancy is considered.

Inspired by Grandvalet and Bengio (2005), the prediction from the source domain usually has low entropy, indicating a highly confident result, while the prediction from the target domain has high entropy, indicating an unreliable result (Vu et al. 2019). Thus, entropy of the resultant map is adopted as the criterion (Shannon 1948). Supposing the Landsat images from source domain (with label) is denoted as \( Y'_m, m = 1 \ldots M \), and the images from target domain (without label) are denoted as \( Y'_l, l = 1 \ldots L \). Given an already-trained model, the output probability map is \( Z'_l \), the entropy of \( Z'_l \) can be given as:

\[
\text{Ent}_{\text{loss}}(Z'_l) = \sum_{i=1}^{H} \sum_{j=1}^{W} \sum_{k=1}^{NC} Z'_l(i, j, k) \cdot \log(Z'_l(i, j, k))
\]

(17)

Generally, enforcing a minimal entropy on the target result \( Z'_l \) can modulate the already-trained network and gradually bridge the domain gap. However, modulating the network parameter takes the risk of negative transferring, which may destroy the performance on the source domain. In order to preserve the sub-pixel-scale classification accuracy on the source domain, we design an alternating optimization entropy minimization (AOE) strategy (Figure 8), to alternately minimize the \( L_{seg}(\hat{Z}_m, Z_m) \) on the source domain and \( \text{Ent}_{\text{loss}}(Z'_l) \) on the target domain:

\[
\hat{\theta} = \text{argmin}\{L_{seg}(\hat{Z}_m, Z_m)\}, m = 1 \ldots M
\]

\[
\hat{\theta} = \text{argmin}\{\text{Ent}_{\text{loss}}(Z'_l)\}, l = 1 \ldots L
\]

(18)

(19)

### 3.3. Experiment setting

#### 3.3.1 Experiment purpose

We firstly validate the proposed DETNet model on the FLAS dataset, together with classical SRM and the state-of-the-art DLSMNet model for comparison, and conduct an ablation analysis to validate the effectiveness of the PESR and AOE. After then, we validate its transferability over large extent of 28 cities in China. Finally, we deploy the DETNet to continuous 21-year Landsat images in Wuhan during 2000–2020, to validate its performance in observing urban dynamic and spatiotemporal evolution patterns. The framework of the experiment is shown in Figure 9.

#### 3.3.2 Evaluation metric

For quantitative validation on FLAS dataset, the per-subpixel classification metrics like overall accuracy (OA), the mean producers’ accuracy (mPA), users’ accuracy (mUA), and mF1-score of all the classes, are investigated. It worth noting that Kappa is recently proved to be improper for land cover accuracy assessment (Foody 2020), and the intersection over union (IoU) essentially measures the same thing as F1-score (Maxwell, Warner, and Andrés Guillén 2021), thus, these two metrics are not included. Meanwhile, in order to examine the performance on each land cover class, class-wise F1-score is also considered. They can be derived from the confusion matrix:

\[
F_{\text{1-score}}(c) = \frac{TP_c}{TP_c + FP_c + FN_c}
\]

(20)

where \( TP_c \) denotes the correct target, \( FP_c \) denotes the false positive, and \( FN_c \) denotes the false negative of the class \( c \).
matrix (Table 2), and the calculation is provided in Eq. (20)-(22).

where \( n_{ij} \) means the number of pixels that belong to class \( j \) (according to the ground truth) is predicted as class \( i \) by the classification model, \( n_{+j} \) represents the total number of pixels that belong to class \( j \), and \( n_{++} \) represents the total number of pixels that are predicted as class \( i \). \( K \) is the total number of classes.

\[
\begin{align*}
\text{mPA} &= \frac{1}{K} \sum_{k=1}^{K} \frac{n_{kk}}{n_{+k}} \quad (20) \\
\text{mUA} &= \frac{1}{K} \sum_{k=1}^{K} \frac{n_{k+}}{n_{++}} \quad (21) \\
\text{mF1 score} &= \frac{2}{K} \sum_{k=1}^{K} \frac{\sum_{k=1}^{K} \frac{n_{kk}}{n_{+k}} \cdot \sum_{k=1}^{K} \frac{n_{kk}}{n_{++}}}{\sum_{k=1}^{K} \frac{n_{kk}}{n_{+k}} + \sum_{k=1}^{K} \frac{n_{kk}}{n_{++}}} \quad (22)
\end{align*}
\]

Furthermore, considering that per-subpixel classification metric only concerns the universally pixel-wise accuracy and cannot reflect the detail reconstruction ability, a novel semivariogram metric is adopted to further inspect the performance on detail structure, which can quantitatively characterize the spatial variation or spatial pattern in land cover map (Curran 1988; Balaguer-Beser et al. 2013). Specifically, after deriving the estimated HR probability map \( \hat{Z} \in \mathbb{R}^{H \times W \times NC} \), the semivariogram at lag \( h \) is:

\[
\gamma_k(h) = \sum_{i=1}^{N(h)} [\hat{Z}_k(v_i) - \hat{Z}_k(v_i + h)], k = 1, \ldots, NC
\]

where \( v_i \) and \( v_i + h \) denote the spatial locations on \( \hat{Z}_k \) that have a distance of \( h \). \( \hat{Z}_k(v_i) \) and \( \hat{Z}_k(v_i + h) \) denote the probability value of these two locations. \( N(h) \) denotes total number of pairs of \( v_i \) and \( (v_i + h) \) on \( \hat{Z}_k \). Specifically, four directions (0°, 45°, 90°, 135°) are considered for lag \( h \), and \( h \) is set to \([1,8]\) with step size 1 in 0° and 90° direction, and with step size 1.414 in 45° and 135° direction (Wang et al. 2020). With the lag varying, the \( \gamma_k(h) \) can form a semivariogram curve in each direction. The more similar the semivariogram curve between the predicted \( \hat{Z} \) and the ground truth \( Z \), the better the performance on the detailed

### Table 2. Confusion matrix.

| Prediction | Class 1 | Class 2 | ... | Class K | Row_sum | UA |
|------------|---------|---------|-----|---------|---------|----|
| Class 1    | \( n_{11} \) | \( n_{12} \) | ... | \( n_{1K} \) | \( n_{+1} \) | \( n_{11}/n_{+1} \) |
| Class 2    | \( n_{21} \) | \( n_{22} \) | ... | \( n_{2K} \) | \( n_{+2} \) | \( n_{22}/n_{+2} \) |
| ...        | ...     | ...     | ... | ...     | ...     | ... |
| Class K    | \( n_{K1} \) | \( n_{K2} \) | ... | \( n_{KK} \) | \( n_{+K} \) | \( n_{KK}/n_{+K} \) |
| Col_sum    | \( n_{+1} = \sum_{i=1}^{K} n_{ii} \) | \( n_{+2} = \sum_{i=1}^{K} n_{i2} \) | ... | \( n_{+K} = \sum_{i=1}^{K} n_{iK} \) | \( n \) | - |
| PA         | \( n_{11}/n_{+1} \) | \( n_{22}/n_{+2} \) | ... | \( n_{KK}/n_{+K} \) | - | - |
structure. Furthermore, in order to quantify this similarity, the mean square difference (MSD) of the semi-variogram curve for each class \( MSD_k \) is investigated.

For quantitative validation on large extent 28 cities in China, three verification strategies are employed: 1) In order to quantitatively validate our products, a consistence assessment with contemporary land cover products is developed. It uses a consistence map derived from other products, including 10-m Esri Land Cover in 2020, 10-m FROMGLC10 in 2017, and 30-m GlobeLand30 in 2020. They are initially resampled to 4-m resolution. If one pixel is recognized as the same class by three products, it is assumed to be a convincing annotation and can be used to validate our product. Although the consistence map is much coarser than our products, it can be still used to evaluate the homogenous area of our products; 2) Considering that the consistence assessment could only validate homogenous area, we further used the publicly certified 4-m GID annotation dataset for further validation. Other prevalent products, including GlobeLand30, FROMGLC10, and Esri Land Cover, are also participated for comparison. Each GID scene has size of \( 6800 \times 7200 \) pixels (~786.36 km\(^2\)). Noteworthy is that GID annotation was essentially random-sampled, which is not always located within the provincial capital city, thus we only use the overlaid GID annotation; 3) Considering that GID annotation dataset is not fine-grained enough to examine the spatial detail reconstruction performance, additional 12 validation regions with size of \( 2345 \times 2345 \) (~88 km\(^2\)) are further sampled in three typical cities, i.e. Beijing, Guangzhou, and Wuhan, at the heterogeneous locations, which are annotated by experts without participating in the experiment. Also, GlobeLand30, FROMGLC10, and Esri Land Cover products are selected for comparison.

### 3.3.3 Comparison algorithms
In order to provide benchmarks for the validation of the proposed method, the classical SRM and the state-of-the-art DLSMNet are considered for comparison. Noteworthy is that since our goal is to super-resolve the coarse Landsat image to fine land cover product, the semantic segmentation network which only deal with the same resolution between input and output is not included. Classical SRM algorithms include spatial attraction SRM model (SASM) (Mertens et al. 2006) and Adaptive MAP model and a winner-takes-all Class Determination strategy for Sub-pixel Mapping (AMCDSM) (Zhong et al. 2015). DLSMNet include LinkNet (Arun, Buddhiraju, and Porwal 2018), SRMCNN (Jia et al. 2019), SPMCNN-ESPCN (He et al. 2021a), SPMCNN-SRGAN (He et al. 2021a), and SIMNet (He et al. 2021b). In addition, traditional pixel-scale classification (HC) is also taken for comparison, which adopted typical SVM method that assigns a single label to each mixed pixel.

### 3.3.4 Experiment environment and parameter settings
All the DLSMNet models are trained on a server with NVIDIA RTX 2080Ti graphics processing unit (GPU) accelerators (11-GB memory) and were implemented in TensorFlow 1.15.0, Pytorch 1.2.0, and Python 3.7.

---

**Figure 9.** Validation experiment framework in this study.
For the state-of-the-art DLSMNet, the epoch is uniformly set to 1000, the learning rate is 0.0001, the Adam optimizer is used, and the batch size is 16, as suggested previously (Arun, Buddhiraju, and Porwal 2018; Jia et al. 2019; He et al. 2021a). As for DETNet, the learning rate is set to 0.01, the batch size is 4, and the SGD optimizer is used with momentum of 0.9 and weight decay of 0.0001. The number of the sampled hard points \( N \) is empirically set to 32,768 (1/8 of the 512 \( \times \) 512 patch) after parameter tuning.

4. Experiment result and validation

4.1. Results on FLAS dataset

4.1.1 Evaluation in terms of per-subpixel classification

Figure 10 demonstrates the visual inspection of the results (five representative resultant patches are selected. More resultant patches can be referred to Appendix S1 in the supplementary file), and Figure 11 displays the zoomed-in region for detail reconstruction examination.

It is obvious that 30 m/pixel Landsat images are affected by the mixed pixel problem, with blurry edges and vague structures, compared to the 1-m Google Earth images. Due to this reason, the result of traditional pixel-scale HC is unpleasant with spackle noise and jagged boundaries. By contrast, sub-pixel-scale classification can generate smooth edges with small-scale land cover objects (Figure 11).

Among compared algorithms, DETNet achieves the best result, with the spatial structure of the building and small-sized grass clearly characterized in 3rd patch, and the elongated road in 2nd patch is also well restored (Figure 10). By contrast, the results of SASM and AMCDSM are compromised with noise, which is attributed to the uncertainty of the input abundance image due to the spectral variability phenomenon. As for DLSMNet, the LinkNet result is over-smoothed without details, since the excessive downsampling in the encoder stage ruins the low-level features. The SRMCNN makes a simplification on LinkNet architecture to reduce pooling layer for preserving spatial details, and the results become better. SRMCNN_ESPCN and SRMCNN_SRGAN eliminate the pooling layer to further keep the low-level feature and use SPC layer to advance the upsampling performance. The SIMNet resembles the structure of SRGAN, further takes the abundance images as semantic constraint for spatial detail enhancement, and achieves a more pleasant result.

The confusion matrices of each result are displayed in Figure 12, with the PA and UA of each class presented. The rows represent the prediction and the columns represent the ground truth. The cell entry is the population proportion of area normalized by dividing the total area, and the values of each table as a whole sum to 1.0 (Stehman and Foody 2019). For DETNet, it is obvious that most predictions are located at the diagonal, and most of the PA and UA of each class are better than other algorithms, indicating that each land cover class is properly predicted. For other algorithms, the shrub and forest are usually confused, as well as the bare land and impervious.

For quantitative assessment in terms of per-subpixel classification metrics (Table 3), the OA of DETNet achieves 79.28%, and outperforms the second best SIMNet by 6.70%. The \( \text{mPA}, \text{mUA} \), and \( mF1 \)-score of DETNet also makes an improvement by 0.0913, 0.1502, and 0.1157 compared to SIMNet, respectively. In addition, the statistical results of SRMCNN, SRMCNN_ESPCN, and SRMCNN_SRGAN are very similar, with average OA of 68.81% and small standard deviation of 0.53%, while LinkNet is slightly worse. Not surprising is that HC ranks the lowest in all accuracy metrics, which indicates an unsatisfactory performance.

For quantitative assessment of each class in terms of \( F1 \)-score (Table 4), the most increment occurred in grass, shrub, and bare land, which are increased by 18.90%, 17.36%, and 23.13%, respectively. The reason is that these classes are sparsely distributed and difficult to be identified for traditional algorithms, while being beneficial to the detail edge localization ability of DETNet, they can be well retrieved. Meanwhile, for cropland, forest, water, and impervious class, DETNet also makes an improvement by 8.43%, 5.21%, 4.36%, and 3.40%, respectively.

4.1.2 Evaluation in terms of semivariogram

The semivariogram curves of the ground truth and that of the predicted results by each algorithm are shown in Figure 13. It is obvious that the curves are similar in the four directions, which indicates a similar spatial variation along these directions. Meanwhile, the curves of each class are varying,
which indicates a largely distinct distribution pattern of each class. For forest, grass, impervious, and water classes, large semivariogram values are observed, indicating that they have distinct distribution patterns with large spatial variation, which is consistent with the visual inspection in Figure 13. For these classes, the DETNet can realize a more similar semivariogram curves with the ground truth, suggesting that DETNet has a superior sub-pixel-scale reconstruction ability.
However, the DETNet has a slightly performance drop for the shrub and bare land, the reason is that the shrub and bare land is rare, and the parcel of these classes are widely spaced and far away from each other compared to the dominant classes, resulting in a low value of the semivariance of the rare classes. Besides, the DETNet is also inferior to SIMNet in restoring the spatial structure of cropland, since the cropland is usually homogeneously distributed, and the structure derived by DETNet tends to be fragmented.

In order to quantify the spatial structure restoration performance, the MSD between the semivariogram curves of each algorithm and that of the ground truth for all the classes are recorded (Table 5). It is obvious that DETNet achieves the best MSD for the forest, grass, water, and impervious classes, which are reduced by 0.0034, 0.0021, 0.0003, and 0.0039, compared to the second best algorithm, respectively. As for cropland, shrub, and bare land, however, the DETNet undergoes accuracy drops of 0.0005, 0.0005, and 0.000068.

4.1.3 Ablation analysis of PESR and AOE

In order to validate the effectiveness of the PESR and AOE strategy, ablation analysis is conducted. The baseline for comparison is constructed by removing both parts. Then, the “Baseline + PESR” is involved by adding PESR, to independently inspect its effectiveness. Finally, the “Baseline + PESR + AOE” (i.e. DETNet) is involved to examine the enhancement achieved by AOE strategy. The statistical result is given in Table 6. Besides, a visual inspection of ablation analysis is presented in Figure 14.

Specifically, PESR makes a remarkable improvement by 8.60% and 0.1458 in OA and mIoU when comparing the “Baseline” with the “Baseline + PESR.” From the visual inspection in Figure 14(a), the accuracy improvement achieved by PESR is mainly located at the boundaries that constitute the detail spatial structure of the scene, which verifies the spatial detail reconstruction ability of PESR.

As for the effectiveness of AOE strategy, the accuracy gain is not obvious, only improved by 0.51% and 0.0068 in terms of OA and mIoU, since
Figure 12. Confusion matrix of each result.

Table 3. Quantitative assessment of each algorithm on FLAS dataset.

| Algorithm   | OA(%) | mPA  | mUA  | mF1-score |
|-------------|-------|------|------|-----------|
| HC          | 40.70 | 0.3263 | 0.3039 | 0.2925 |
| DeepLabv3+  | 65.83 | 0.4925 | 0.3687 | 0.4925 |
| SASM        | 42.61 | 0.3867 | 0.3529 | 0.3136 |
| AMCDSM      | 45.14 | 0.4048 | 0.3672 | 0.3316 |
| LinkNet     | 62.51 | 0.4143 | 0.4229 | 0.4046 |
| SRMCNN      | 68.78 | 0.5189 | 0.5460 | 0.5145 |
| SRMCNN_EPSCN| 69.47 | 0.4972 | 0.5823 | 0.5024 |
| SRMCNN_SRGAN| 68.17 | 0.4797 | 0.5333 | 0.4845 |
| SIMNet      | 72.58 | 0.5922 | 0.6375 | 0.6040 |
| DETNet      | 79.28 | 0.6835 | 0.7877 | 0.7197 |

Table 4. The F1-score (%) accuracy for each class by different algorithms.

| Algorithm | Cropland | Forest | Grass | Shrub | Water | Impervious | Bare land |
|-----------|----------|--------|-------|-------|-------|------------|-----------|
| HC        | 34.18    | 51.05  | 8.72  | 5.02  | 44.50 | 95.04      | 8.95      |
| SASM      | 29.82    | 54.55  | 5.90  | 15.47 | 67.54 | 58.66      | 8.95      |
| AMCDSM    | 30.73    | 57.23  | 6.07  | 14.96 | 75.07 | 61.70      | 8.87      |
| LinkNet   | 47.06    | 70.53  | 8.14  | 14.96 | 74.98 | 74.98      | 8.95      |
| SRMCNN    | 63.02    | 76.15  | 17.48 | 30.53 | 74.92 | 78.11      | 19.97     |
| SRMCNN_EPSCN | 62.08 | 76.01  | 24.44 | 20.09 | 75.78 | 79.34      | 13.96     |
| SRMCNN_SRGAN | 59.39 | 75.11  | 21.17 | 14.01 | 75.07 | 78.58      | 15.84     |
| SIMNet    | 68.83    | 78.58  | 35.50 | 47.53 | 77.07 | 80.38      | 35.09     |
| DETNet    | 77.26    | 83.79  | 54.40 | 64.89 | 81.43 | 83.78      | 58.22     |
Figure 13. The semivariogram curves of each algorithm to examine the reconstruction ability toward the spatial detail structure.

Table 5. The MSD between the semivariogram curves of each algorithm and that of ground truth for each class.

| Algorithm   | Cropland | Forest  | Grass   | Shrub   | Water   | Impervious | Bare land |
|-------------|----------|---------|---------|---------|---------|------------|-----------|
| LinkNet     | 0.0018   | 0.0345  | 0.0227  | 0.0017  | 0.0057  | 0.0396     | 0.0004    |
| SRMCNN      | **0.0005** | 0.0279  | 0.0201  | 0.0012  | 0.0035  | 0.0267     | **0.0003** |
| EPSCN-SRMCNN | 0.0033  | 0.0166  | 0.0123  | 0.0017  | 0.0025  | 0.0214     | 0.0001    |
| SRGAN-SRMCNN | 0.0007  | 0.0239  | 0.0180  | 0.0016  | 0.0038  | 0.0252     | 0.0002    |
| SIMNet      | 0.0010   | 0.0114  | 0.0088  | **0.0000** | 0.0018  | 0.0152     | 0.00009   |
| DETNet      | 0.0010   | **0.0080** | **0.0067** | 0.0005  | **0.0011** | **0.0113** | 0.0001    |
the statistical distribution of the training set and test set of FLAS dataset is similar. Despite this, from the visual inspection in Figure 14(b), the cropland and grass land, which is easily confused in “Baseline + PESR,” is properly distinguished in “Baseline + PESR + AOE,” owing to the contribution of AOE strategy.

### 4.2 Results on large-extent study area of the 28 cities in China

The large-scale mapping products of the 28 cities in China in 2020 derived by DETNet are summarized in Figure 15. A detailed visual inspection of them can be found in Appendix S2 in the supplementary file. Besides, the mapping results on the residual six provincial capital cities located at the continental climate and mountain climate zone are separately provided in Appendix S3, to provide the limitation of our model.

#### 4.2.1 Evaluation by contemporary land cover products

From consistent assessment (provided in Appendix S4 in the supplementary file), it is obvious that the accuracy of the cities in China’s north-eastern region (e.g. Shenyang, Harbin, Changchun, Beijing, and Hefei) is usually higher than that in China’s central plain and south region (e.g. Wuhan, Changsha, Chongqing, and Haikou), since the land covers in the north cities are more homogenous cropland or forest gathered to a cluster, while they are scatteredly distributed in the south cities with large spatial heterogeneity. The highest consistency of 95.06% OA is achieved in Shenyang, which is due to its flat terrain with clear boundaries between urban impervious area and cropland. Besides, the cropland is homogeneously distributed with little forest or water disturbance. While the lowest consistency of 74.68% OA is achieved in Chongqing, which due to its mountainous topographic makes the urban and cropland highly confused with the mountain forest. Overall, our products achieve well consistency with the prevailing contemporary products, achieving OA of 84.11%, which preliminarily substantiate its reliability. Noteworthy is that the consistence map only has values at the homogenous area, thus the accuracy presented can only reflect the rough performance at the homogenous area, further examination at heterogenous area should be considered.

#### 4.2.2 Evaluation by GID annotation

The visual and quantitative examination validated by the GID annotation are provided in Appendix S2 and S5, respectively, which has totally 14 cities.

#### Table 6. Ablation analysis for the validation of PESR and AOE strategy based on FLAS dataset, in terms of OA and mIoU.

| Method                     | PESR | AOE | Loss               | OA (%) | mIoU  |
|----------------------------|------|-----|--------------------|--------|-------|
| Baseline                   |      |     | SRM_loss           | 70.33  | 0.4174|
| Baseline+PESR              | ✓    |     | SRM_loss+Point_loss| 78.93  | 0.5632|
| Baseline+PESR+EM           | ✓    | ✓   | SRM_loss+Point_loss+Ent_loss| 79.44  | 0.5700|

**Figure 14.** The visual inspection of the improvement achieved by the PESR and the AOE strategy. (a) Comparison between “Baseline” and “Baseline + PESR”; (b) comparison between “Baseline + PESR” and “Baseline + PESR + AOE.”

**Figure 15.** Ablation analysis for the validation of PESR and AOE strategy based on FLAS dataset, in terms of OA and mIoU.
that have overlaid region with our products. From visual inspection, it can be observed that the GlobeLand30 product is relatively coarse, the urban area is usually clustered to one polygon without inner structure, which is not suitable for observing urban dynamic. Although at 10-m resolution, the Esri Land Cover product is also coarse, and the cropland area is usually misclassified to impervious surface. As for FROMGLC10 products, the spatial heterogeneity is well delineated; however, the cropland at outskirt is usually misclassified to grass land. On the contrary, the DETNet-derived products restore an urban structure with rich details and fine-grained boundaries.

For quantitative assessment, our products achieve the best accuracy for most of the cities. It worth noting that the GID is annotated on Gaofen-2 images in 2015–2016, which is not exactly matched the date of our products in 2020, resulting in an accuracy fallen in Hangzhou and Chengdu, which experienced a rapid urbanization. Averagely, our products achieve OA of 83.11% and surpass the state-of-the-art Esri Land Cover products by 8.515%.

4.2.3 Evaluation on three typical cities

From visual examination in Wuhan area (Figure 16), the detailed spatial structure is well restored in the DETNet result. For example, in T1 region located at the optical industrial park center, the detailed patterns with intersection of the cropland, grass, as well as the district blocks and the road nets, are distinct in DETNet result. As in T2 region located at the Huazhong Agricultural University in downtown area, the structural buildings, urban forest, and cropland are clearly distinguished in DETNet result, but are aggregated in buildings without any structure patterns in other compared products. The visual examination and analysis in Beijing and Guangzhou are provided in Appendix S6.

For quantitative assessment (Table 7), the result achieved by DETNet is superior to Esri Land Cover, FROMGLC10, and GlobeLand30 product in the three cities. Averagely, DETNet achieves OA of 70.70% and outperforms the second-best product by 1.28%. Combined with visual analysis, the increments are mainly located at the heterogeneous region, in which the other products only characterize a homogenous pattern without
detailed structure, causing an accuracy drop. Although FROMGLC10 reveals considerable spatial details, the recognition is not accurate with large amount of grass area.

4.2.4 Ablation analysis of AOE

Due to the various imaging conditions, the statistical distribution of the training set has considerable distinction in the Landsat images over large extent. In

Figure 16. Demonstration of the land cover products in Wuhan.
order to validate the effectiveness of the AOE strategy in generalizing the model to the large extent and quantify its contribution, an ablation analysis is conducted, by comparing the result of “Baseline + PESR” and “Baseline + PESR + AOE,” at Beijing, Guangzhou, and Wuhan areas. Figure 17 demonstrates a visual perceptual of the contribution of the AOE strategy, and Table 8 provides an average quantitative assessment on the validation regions used in section 4.2.3.

From visual inspection, the results achieved by “Baseline + PESR” are affected by noise, which is due to the fact that the trained model cannot properly recognize the unknown scene that has different statistical distribution from the training set and only gives ambiguous probability maps with a variety of uncertainty points. By contrary, the result with AOE strategy is more continuous, since the noise is appropriately surpassed by reducing the entropy of the target result to transform it to the manifold of the source domain, which can alleviate the uncertainty in probability maps of the target domain.

### Table 7. Quantitative assessment of each product at validation regions in Beijing, Guangzhou, and Wuhan cities.

| Area   | Methods   | OA (%) | mPA    | mUA    | mF1-score |
|--------|-----------|--------|--------|--------|-----------|
| Beijing| GlobeLand30| 70.62  | 0.4463 | 0.5318 | 0.4535    |
|        | FROMGLC  | 65.72  | 0.4450 | 0.6706 | 0.4605    |
|        | Esri     | 72.87  | 0.5019 |        | 0.7086    |
|        | DETNet   | **73.59** | **0.5434** | 0.5839 | **0.5440** |
| Guangzhou| GlobeLand30| 71.32  | 0.5643 | 0.5601 | 0.5426    |
|         | FROMGLC  | 51.35  | 0.5240 | 0.6395 | 0.5310    |
|         | Esri     | 61.56  | 0.5865 | 0.5808 | 0.5264    |
|         | DETNet   | **71.36** | **0.6665** | **0.6411** | **0.6286** |
| Wuhan  | GlobeLand30| 66.30  | 0.3330 | 0.3890 | 0.3337    |
|        | FROMGLC  | 62.80  | **0.4208** | **0.4649** | 0.3982    |
|        | Esri     | 67.03  | 0.3976 | 0.4416 | 0.3742    |
|        | DETNet   | **67.15** | 0.3971  | 0.4216 | **0.3992** |

### Table 8. Ablation experiment for the validation of the AOE strategy on Beijing, Wuhan, and Guangzhou areas.

| Area   | Methods   | OA (%) | mPA    | mUA    | mF1-score |
|--------|-----------|--------|--------|--------|-----------|
| Beijing| Without AOE| 62.62  | **0.5717** | 0.5097 | 0.4939    |
|        | With AOE  | **73.59** | 0.5434  | **0.5839** | **0.5440** |
| Guangzhou| Without AOE| 63.36  | 0.6254 | 0.6194 | 0.5684    |
|         | With AOE  | **71.36** | **0.6665** | **0.6411** | **0.6286** |
| Wuhan  | Without AOE| 62.33  | **0.4280** | 0.3966 | 0.3875    |
|        | With AOE  | **67.15** | 0.3971  | **0.4216** | **0.3992** |
From quantitative assessment, the AOE strategy significantly enhances the classification accuracy, by 10.97%, 8.00%, and 4.82% in terms of OA in Beijing, Guangzhou, and Wuhan, respectively. This improvement indicates that a considerable cross-domain discrepancy was filled up by the AOE strategy and confirms the transferability of our method for a large spatial extent.

4.3 Results on long-term and fine-scale land cover mapping of the 21 years in Wuhan

In addition to spatial extent, the DETNet is also deployed to long temporal extent of continuous 21 years (2000–2020) at Wuhan city that experienced a rapid and constant urbanization, to examine its transferability in capturing fine-grained urban dynamic. Specifically, Landsat-5 within 2000–2011, Landsat-7 in 2012, and Landsat-8 within 2013–2020 are used to form a series of 21-year observations. A median mosaic of the growing period ranging from April to September in each year is taken to eliminate the cloud contamination and phenology variation. Besides, in order to ensure consistency over time, a histogram matching method (Wegener 1990) is initially employed to adjust Landsat-5, 7, and 8 to have the same mean and variance, and 2020 is selected as the reference for the matching. Furthermore, a continuous change detection (CCD) method (Zhu and Woodcock 2014) is performed on annual time-series products predicted by DETNet, to relieve the fluctuation of possible uncertainty. Specifically, we only use the break map (indicating which pixel is varied in which year) in CCD as the

Figure 18. The spatiotemporal change patterns of Tianhe Airport in Wuhan, China.
indicator mask for the refinement of our time-series mapping results, which can eliminate noise and false changes. The annual time-series results are provided in Figure 18, and a zoomed-in area at Tianhe Airport is provided for demonstrating the performance on detail urban composition. Noteworthy is that other land cover products are limited to the production period and only have single or several years’ results, which is not suitable for studying the spatiotemporal change of the surface process.

From visual inspection of the fine-grained urban composition, the morphological structure of Airport Terminal-1 is clearly reconstructed in 2000, such as the lawn and track in the airport. Besides, the construction of Airport Terminal-2 was begun in 2004 and completed in 2008, Airport Terminal-3 was begun in 2012 and completed in 2016, the spatiotemporal dynamic of which is consistent with our time-series products, proving the ability of our products in observing fine-grained urban dynamic.

From the quantitative statistic of the land cover area variation (annual variation of each land cover displayed in Figure 19(a) and the growing increment every 5 years displayed in Figure 19(b)), the impervious area is stably growing with average speed of 16.87 km², except for 2015–2020, which experienced a prosperity of the real estate market that makes the housing prices in Wuhan almost doubled. From the dynamic maps in Figure 20, it is obvious that the impervious area gradually encroaches the cropland along the outskirt. Although being encroached, the cropland area is still constantly grown, since it instead encroaches the lake and forest area, as can be seen in Figure 20. The growing rate of the cropland area fluctuated with an average speed of 29.71 km². Especially in 2015–2020, cropland experiences a drastic increase, the reason behind it is partially attributed to the new first-tier city of Wuhan and the increasing population that put considerable pressure on the agriculture yield. Besides, Wuhan, Hubei is also the main producing location for paddy rice in central plain of China.

From urban structure dynamic reflected in three main districts, i.e. Wuchang, Hanyang, and Hankou (zoomed-in every 5 years in Figure 21). We find that the impervious areas were separately distributed in the old city zone and outer city zone in early year, with the village region abruptly located between them. This is the urban-village phenomenon in
China after “Reform and Open-up” policy, which was caused by rapid urbanization without rational urban planning that gradually encircled the village with high buildings. As time goes on, the impervious areas were then gradually dispersed from the old city zone to the outer city zone and finally connected, forming the Lingkonggang, Zhuankou, and East Lake High-tech Development Zone, respectively, during which process the middle village regions were disappeared. This is mainly due to the urban-village renovation policy that is aimed to redistribute the spatial configuration of the city to renovate the village and construct a healthy and sustainable urban structure.

4.4 Computational efficiency
In practical land cover mapping activity, the time consumption is one of the critical indicators when
applying DETNet to large spatial extent. In order to examine the computational efficiency of the proposed DETNet, the time consumption during the training phase, the fine-tuning phase, and the inference phase in Wuhan area are recorded, respectively (Table 9). In practical situation, only the fine-tuning phase and the inference phase are needed when new remote sensing image comes up. After statistic, the DETNet model only takes 1788.13 s to infer the Wuhan area with coverage of 8,569 km², which has great potential for achieving frequency updating of the land cover product, and it is of great importance for meeting the requirement of frequency monitoring applications.

5. Discussion

These results show that the nonuniformed sampling that clustered at the low confidence edge position when upsampling feature maps can substantially restore more details than the classical deconvolution or sub-pixel convolution in previous DLSMNet that adopted fixed grid sampling of 3 × 3 (or other size) kernel. This finding is consistency with our hypothesis that the learning of the spatial heterogenous distribution is usually weakened by the learning of the homogenous distribution, which is partly attributed to fact that the fixed grid sampling of convolution operation favors spatial homogeneity. Besides, the results also show that synergistically considering the entropy of the target domain during model fine-tuning can considerably reduce the influence of the geometric or radiance variations between source and target domain, which make it possible to transfer to a large extent application.

Based on the above two advanced properties, the proposed DETNet model is able to break through the spatial and temporal limitations of the current land cover mapping activities, by super-resolution 30-m Landsat images to meter-level fine-grained land cover distribution. Once trained, the DETNet model can be directly deployed on large-extent and multi-temporal Landsat images without human intervention, which costs less and has computational efficiency. The high spatial and temporal resolution land cover product is of great value and indicative two aspects, such as the case in Section 4.3: 1) From the area statistic and spatial variation, the impact of urbanization to the ecosystem is indirect, i.e. imperious surface encroached cropland, while cropland encroached forest. Besides, the growth rate of cropland experienced a large increase during 2015–2020 in Wuhan (Province of a Thousand Lakes), which largely encroached the lake area. 2) From micro view of urban structure and its evolution patterns, the dynamic of the impervious surface in the old city zone and the outer city zone was well captured, which reflected an urban-village phenomenon and urban-village renovation process reforming the urban structure. Therefore, the derived products can be served as a data support for municipal decision-making.

Although the effect of the modification of the network model is obvious, the limitations of the remote sensing dataset and annotations are also nonnegligible. Specifically, the FLAS dataset is only sampled in temperate and subtropical monsoon climate zone and is mainly located around the urban area, which lacks the samples of wetland, tundra, and ice/snow. In this situation, the trained model may be unable to identify these land covers in the mountainous zone of northwest China. Besides, due to the class imbalance issue (i.e. small amount of grass, shrub, and bareland samples), the transferability of the trained model may be limited. Specifically, it is more suitable in temperate monsoon climate zone, the subtropical monsoon climate zone, or urban agglomeration and its surrounding area, than in other area like desert, mountain, or prairie. But noteworthy is that the advantage of our model is the detail reconstruction ability of heterogenous region with significant spatial variation, such as urban area. The class imbalance issue can be also alleviated by reducing the weight of large ratio class and enlarging the weight of small ratio class in the loss function calculation.

Furthermore, satellite image quality is often degraded by image striping due to sensor calibration problems, as well as by thin clouds and fractus clouds, especially in tropical climates with year-round rainfall. These and other issues reduce the usefulness of spectral data for image classification. In this situation, a data fusion method needs to be further explored to restore the missing/contaminated observations.

In the future development of global land cover mapping community, the DETNet model has great potential to realize seamless and fine-grained HR land cover products, which is indicative of the precise
spatial variation of each land covers, can facilitate greater understanding of the impacts of urbanization on biodiversity and human settlements, and help to achieve sustainable development goals.

6. Conclusion

HR global land cover product is a fundamental source for retrieving Earth surface processes and understanding the interaction of urban dynamic and environmental variation. However, current products are usually restrained by the spatial resolution (up to 10 m) and temporal resolution (2000, 2010, 2020 for GlobeLand30; 2010, 2015, 2017 for FROMGLC; 2020 for Esri Land Cover), and the urban composition is highly heterogenous, which make the current products insufficient to understand how cities change in space and over time at HR.

This study attempts to alleviate this major problem in the aspect of mixed pixel problem and use advanced super-resolution mapping (SRM) method to decompose the mixed pixel (30 × 30 m) and locate each land cover at sub-pixel-scale (cira 4 × 4 m), which consequently improves the spatial resolution of the land cover map. Considering two specific limitations in deep-learning-based SRM community, i.e. 1) the fixed grid sampling of 3 × 3 kernel in the deconvolution or sub-pixel convolution in previous DLSTMNet favors spatial homogeneity and suppresses the learning of spatial heterogeneity of urban composition; 2) the geometric or radiation variation in different region and time in Landsat images makes it difficult for simple SRM to transfer, we develop the DETNet model. Two innovations are obvious: 1) PESR architecture is designed to allow nonuniformed sampling specific at the low-confidence edge position, which can intensify the learning of heterogenous compositions’ patterns that are surpassed in the fixed grid sampling of classical deconvolution; 2) AOE minimization strategy is designed to alternatingly minimize the entropy of the target domain and the SRM loss of the source domain during model fine-tuning, to gradually project the classification hyperplane to the manifold of the target domain, making the DETNet model valid on Landsat image in large-extent and different time.

From experimental results on FLAS dataset, we find that DETNet achieves outstanding performance on detail reconstruction in heterogenous region, the OA of which reaches 79.44%, and makes 6.86% accuracy improvement. From ablation analysis, we find that PESR makes an accuracy improvement of 8.60% in OA, while AOE makes an averagely 7.93% promotion when transferring DETNet in Beijing, Guangzhou, and Wuhan. The semivariogram curve criterion also shows that DETNet achieves the most similar spatial variation with the ground truth.

To examine its transferability in large spatial extent, we deploy DETNet to the 28 cities in China and validate the derived products by three quantitative verification strategies. We find that the derived products restore considerable heterogenous distribution in delineating the urban composition and structure and achieve better accuracy than the contemporary Esri Land Cover, FROM-GLC10, and GlobeLand30 products. To examine its transferability in different times, we further deploy it on continuous 21-year Landsat images during 2000–2020 in Wuhan city and find how city changes in space and over time at HR in two aspects: 1) From area statistic of each land cover, it is found that even encroached by impervious surface, the expansion rate of cropland is much faster than the impervious area over the past 10 years, which are the principal source of the encroachment of forest, grass, and lakes; 2) From the spatiotemporal evolution patterns, it is found that anisotropic rapid urbanization makes the old village encircled with new buildings in early years, while these urban-villages are then reconstructed in later years to redistribute the spatial arrangement for a more rational urban structure, after the execution of urban-village renovation policy. Therefore, the derived products are able to provide data support for assisting the municipal decision-making.

Some key issues can be pursued in future to further promote this research and facilitate a seamless national-scale/global-scale high resolution product. Firstly, samples in the continental climate and mountain climate zone can be increased, to enhance the generalization of the network model. Secondly, the bias in land cover types in FLAS annotations may also result in uncertainty in predicting minority class, such as grass and bare land. Thirdly, considering the single source of spatial prior only from training dataset may not be sufficient to constrain the ill-posed SRM problem, the idea of spatio-temporal super-resolution mapping method can be adopted, which borrows the complementary HR spatial patterns from other temporal images. In summary, this work provides a novel SRM method for generating HR
land cover product, and it is the first attempt to transfer the SRM model to the largest extent that has ever been done before, which make it possible for mass production of land cover maps. We hope that the DETNet as well as the derived products could be used as a support for further geographical-related research.

Acknowledgements

This research was performed on NVIDIA RTX 2080Ti graphics processing unit accelerators. We thank the anonymous reviewers for their insightful comments and suggestions, which help improved our manuscript.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the China Postdoctoral Science Foundation [2020M683053]; National Key Research and Development Program of China [2019YFB2103102]; National Natural Science Foundation of China [61976234]; Natural Science Foundation of Guangdong Province [2019A1515011057, 2020A1515110708].

ORCID

Qian Shi http://orcid.org/0000-0002-1276-0352

Data availability statement

The DETNet-derived 28 cities products, the annual 21-year Wuhan products, and the execution code can be downloaded at https://doi.org/10.5281/zenodo.6899503. The products are named by the city’s name. The corresponding comparative products (including Esri Land Cover, FROMGLC10, and GlobeLand30), as well as the validation dataset and validation code for the three verification strategies are also available at this URL.

Author contributions

D. He designed the proposed DETNet network model, which is based on the idea of DLSMNet from Y. Zhong. D. He programmed the code to perform the experiments in section 4.1 under the guidance of Q. Shi. D. He wrote the original draft. Q. Shi provides the FLAS annotation dataset, which is critical for training the network model. Q. Shi provides the annotations in 12 validation regions in three typical cities in section 4.2.3. X. Liu provides the idea of extending the model to the spatial extent of the 28 cities (Section 4.2) and the temporal extent of decades (Section 4.3) to examine the transferability of the proposed method, and provides some discussion about the meaningful dynamic that is reflected by our annual products. L. Zhang provides the idea of ablation analysis in section 4.1.3 and section 4.2.4. G. Xia provides the GID annotation dataset and proposed validating the derived products by GID annotation in section 4.2.2.

References

Adams, J. B., D. E. Sabol, V. Kapos, R. A. Filho, D. A. Roberts, M. O. Smith, and A. R. Gillespie. 1995. “Classification of Multispectral Images Based on Fractions of Endmembers: Application to land-cover Change in the Brazilian Amazon.” Remote Sensing of Environment 52 (2): 137–154. doi:10.1016/0034-4257(94)00098-8.

Aguirre-Gutiérrez, J., S. Rifai, A. Shenkin, I. Oliveras, L. P. Bentley, M. Svátek, C. A. J. Girardin, et al. 2020. “Pantropical Modelling of Canopy Functional Traits Using Sentinel-2 Remote Sensing Data.” Remote Sensing of Environment 252:112122. doi:10.1016/j.rse.2020.112122.

Aljaddani, A. H., X.-P. Song, and Z. Zhu. 2022. “Characterizing the Patterns and Trends of Urban Growth in Saudi Arabia’s 13 Capital Cities Using a Landsat Time Series.” Remote Sensing 14 (10): 2382. doi:10.3390/rs14102382.

Ardila, J. P., V. A. Tolpekin, W. Bijker, and A. Stein. 2011. “Markov-random-field-based super-resolution Mapping for Identification of Urban Trees in VHR Images.” ISPRS Journal of Photogrammetry and Remote Sensing 66 (6): 762–775. doi:10.1016/j.isprsjprs.2011.08.002.

Arun, P. V., K. M. Buddhiraju, and A. Porwal. 2018. “CNN Based Super-resolution Mapping for Hyperspectral Images.” Neurocomputing 311: 51–64. doi:10.1016/j.neucom.2018.05.051.

Atkinson, P. M. 1997. “Mapping Sub-pixel Boundaries from Remotely Sensed Images.” Innovations in GIS IV. London: U. K. Taylor and Francis, 166–180. https://www.taylorfrancis.com/chapters/edit/10.1201/9781482272956-25/mapping-sub-pixel-boundaries-remotely-sensed-images-peter-atkinson

Atkinson, P. M. 2005. “Sub-pixel Target Mapping from Soft-classified, Remotely Sensed Imagery.” Photogrammetric Engineering and Remote Sensing 71 (7): 839–846. doi:10.14358/PERS.71.7.839.

Atkinson, P. M. 2009. “Issues of Uncertainty in Super-resolution Mapping and Their Implications for the Design of an Inter-comparison Study.” International Journal of Remote Sensing 30 (20): 5293–5308. doi:10.1080/01431160903131034.

Balagué-Beser, A., L. A. Ruiz, T. Hermosilla, and J. A. Recio. 2013. “Using Semivariogram Indices to Analyse Heterogeneity in Spatial Patterns in Remotely Sensed Images.” Computers & Geosciences 50: 115–127. doi:10.1016/j.cageo.2012.08.001.

Bartholome, E., and A. S. Belward. 2005. “GLC2000: A New Approach to Global Land Cover Mapping from Earth
Observation Data.” *International Journal of Remote Sensing* 26 (9): 1959–1977. doi:10.1080/01431160412331291297.

Bicheron, P., V. Amberg, L. Bourg, D. Petit, M. Huc, B. Miras, C. Brockmann, et al. 2011. “Geolocation Assessment of MERIS GlobCover Orthorectified Products.” *IEEE Transactions on Geoscience and Remote Sensing* 49 (8): 2972–2982. doi:10.1109/Tgrs.2011.2122337.

Bioucas-Dias, J. M., A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, J. Chanussot, et al. 2012. “Hyperspectral Unmixing Overview: Geometrical, Statistical, and Sparse Regression-Based Approaches.” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 5 (2): 354–379. doi:10.1109/JSTARS.2012.2194696.

Boucher, A., and P. C. Kyriakidis. 2006. “Super-resolution Land Cover Mapping with Indicator Geostatistics.” *Remote Sensing of Environment* 104 (3): 264–282. doi:10.1016/j.rse.2006.04.020.

Burke, M., A. Driscoll, D. B. Lobell, and S. Ermon. 2021. “Using Satellite Imagery to Understand and Promote Sustainable Development.” *Science* 371 (6535): 1219. doi:10.1126/science.abe8628.

Chen, J., Y. Ban, and S. Li. 2014. “China: Open Access to Earth land-cover Map.” *Nature* 514 (7523): 434. doi:10.1038/514434a.

Chen, Y., Y. Ge, R. An, and Y. Chen. 2018b. “Super-Resolution Mapping of Impervious Surfaces from Remotely Sensed Imagery with Points-of-Interest.” *Remote Sensing* 10 (2): 242. doi:10.3390/rs10020242.

Chen, L., G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille. 2018a. “DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs.” *IEEE Transactions on Pattern Analysis and Machine Intelligence* 40 (4): 834–848. doi:10.1109/TPAMI.2017.2699184.

Chen, Y., K. Shi, and Y. Ge. 2021b. “Spatiotemporal Remote Sensing Image Fusion Using Multiscale Two-Stream Convolutional Neural Networks.” *IEEE Transactions on Geoscience and Remote Sensing* 60: 4402112. doi:10.1109/TGRS.2021.3069116.

Chen, Y., Q. Weng, L. Tang, Q. Liu, X. Zhang, and M. Bilal. 2021a. “Automatic Mapping of Urban Green Spaces Using a Geospatial Neural Network.” *GIScience & Remote Sensing* 58 (4): 624–642. doi:10.1080/15481603.2021.1933367.

Curran, P. 1988. “The semi-variogram in Remote Sensing: An Introduction.” *Remote Sensing of Environment* 24 (3): 493–507. doi:10.1016/0034-4257(88)90021-1.

Dai, S., M. Han, W. Xu, Y. Wu, and Y. Gong. 2007. “Soft Edge Smoothness Prior for Alpha Channel Super Resolution.” *IEEE Conference on Computer Vision and Pattern Recognition*, Minneapolis, MN, USA, 1–8. doi:10.1109/CVPR.2007.383028.

Deng, C., C. Li, Z. Zhu, W. Lin, and L. Xi. 2017. “Subpixel Urban Impervious Surface Mapping: The Impact of Input Landsat Images.” *ISPRS Journal of Photogrammetry and Remote Sensing* 133: 89–103. doi:10.1016/j.isprsjprs.2017.09.015.

Deng, C., and Z. Zhu. 2020. “Continuous Subpixel Monitoring of Urban Impervious Surface Using Landsat Time Series.” *Remote Sensing of Environment* 238: 110929. doi:10.1016/j.rse.2018.10.011.

Fisher, P. 1997. “The Pixel: A Snare and A Delusion.” *International Journal of Remote Sensing* 18 (3): 679–685. doi:10.1080/014311697219015.

Foody, G. M. 2020. “Explaining the Unsuitability of the Kappa Coefficient in the Assessment and Comparison of the Accuracy of Thematic Maps Obtained by Image Classification.” *Remote Sensing of Environment* 239: 111630. doi:10.1016/j.rse.2019.111630.

Friedl, M. A., D. K. McVeer, J. C. F. Hodges, X. Y. Zhang, D. Muchoney, A. H. Strahler, C. E. Woodcock, et al. 2002. “Global Land Cover Mapping from MODIS: Algorithms and Early Results.” *Remote Sensing of Environment* 83 (1–2): 287–302. doi:10.1016/S0034-4257(02)00078-0.

Fry, J., G. Xian, S. Jin, J. Dewitz, C. Homer, L. Yang, C. Barnes, N. Herold, and J. Wickham. 2011. “Completion of the 2006 National Land Cover Database for the Conterminous United States.” *Photogrammetric Engineering and Remote Sensing* 77 (9): 858–864.

Ge, Y. 2013. “Sub-pixel land-cover Mapping with Improved Fraction Images upon multiple-point Simulation.” *International Journal of Applied Earth Observation and Geoinformation* 22: 115–126. doi:10.1016/j.jag.2012.04.013.

Ge, Y., Y. Chen, A. Stein, S. Li, and J. Hu. 2016. “Enhanced Subpixel Mapping with Spatial Distribution Patterns of Geographical Objects.” *IEEE Transactions on Geoscience and Remote Sensing* 54 (4): 2356–2370. doi:10.1109/TGRS.2015.2499790.

Gerke, M., F. Rottensteiner, J. D. Wegner, and G. Sohn. 2014. “Ispsr Semantic Labeling Contest. In: Photogrammetric Computer Vision.” https://www2.isprs.org/commissions/comm3/wg4/semantic-labeling.html

Gong, P., J. Wang, L. Yu, Y. C. Zhao, Y. Y. Zhao, L. Liang, Z. Niu, et al. 2013. “Finer Resolution Observation and Monitoring of Global Land Cover: First Mapping Results with Landsat TM and ETM+ Data.” *International Journal of Remote Sensing* 34 (7): 2607–2654. doi:10.1080/01431161.2012.748992.

Grandvalet, Y., and Y. Bengio. 2005. “Semi-supervised Learning by Entropy Minimization.” *Advances in Neural Information Processing Systems (NIPS)*. https://proceedings.neurips.cc/paper/2004/file/96f2b50b5d3613adfa9c27049b2a888c7-Paper.pdf

He, D., Q. Shi, X. Liu, Y. Zhong, and X. Liu. 2021c. “Spectral-spatial Fused Sub-pixel Mapping Based on Deep Neural Network.” *IEEE Geoscience Remote Sensing Letter* 19: 1–5. doi:10.1109/LGRS.2021.3101499.

He, D., Q. Shi, X. Liu, Y. Zhong, and X. Zhang. 2021b. “Deep Subpixel Mapping Based on Semantic Information Modulated Network for Urban Land Use Mapping.” *IEEE Transactions on Geoscience and Remote Sensing* 59 (12): 10628–10646. doi:10.1109/TGRS.2021.3050824.

He, D., Q. Shi, X. Liu, Y. Zhong, and L. Zhang. 2022. “Generating 2m fine-scale Urban Tree Cover Product over 34 Metropolises in China Based on Deep context-aware sub-pixel Mapping Network.” *International Journal of Applied Earth Observation and Geoinformation* 106: 102667. doi:10.1016/j.jag.2021.102667.

He, K., X. Zhang, S. Ren, and J. Sun. 2016b. “Deep Residual Learning for Image Recognition.” *IEEE Conference on*
Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 770–778. doi: 10.1109/CVPR.2016.90.

He, D., Y. Zhong, R. Feng, and L. Zhang. 2016a. “Spatial-Temporal Sub-Pixel Mapping Based on Swarm Intelligence Theory.” Remote Sensing 8 (11): 30. doi: 10.3390/rs8110894.

He, D., Y. Zhong, X. Wang, and L. Zhang. 2021a. “Deep Convolutional Neural Network Framework for Subpixel Mapping.” IEEE Transactions on Geoscience and Remote Sensing 59 (11): 9518–9539. doi: 10.1109/TGRS.2020.3032475.

Huan, L., N. Xue, Z. Zheng, W. He, J. Gong, and G. Xia. 2022. “Unmixing Convolutional Features for Crisp Edge Detection.” IEEE Transactions on Pattern Analysis and Machine Intelligence 44 (10): 6602–6609. doi: 10.1109/TPAMI.2021.3084197.

Hu, X., P. Zhang, Q. Zhang, and J. Wang. 2021. “Improving Wetland Cover Classification Using Artificial Neural Networks with Ensemble Techniques.” GIScience & Remote Sensing 58 (4): 603–623. doi: 10.1080/15481603.2021.1932126.

Jia, Y., Y. Ge, Y. Chen, S. Li, G. B. Heuvelink, and L. Feng. 2019. “Super-Resolution Land Cover Mapping Based on the Convolutional Neural Network.” Remote Sensing 11 (15): 1815. doi:10.3390/rs11151815.

Karra, K., C. Kontgis, Z. Statman-Well, J. C. Mazzariello, M. Mathis, and S. P. Brumby. 2021. “Global Land use/land Cover with Sentinel 2 and Deep Learning.” IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Brussels, Belgium, 4704–4707. doi: 10.1109/IGARSS47720.2021.9553499.

Kasetkasem, T., M. K. Arora, and P. K. Varshney. 2005. “Super-resolution Land Cover Mapping Using a Markov Random Field Based Approach.” Remote Sensing of Environment 96 (3–4): 302–314. doi:10.1016/j.rse.2005.02.006.

Keshava, N., and J. F. Mustard. 2002. “Spectral Unmixing.” IEEE Signal Processing Magazine 19 (1): 44–57. doi: 10.1109/79.974727.

Kipf, T. N., and M. Welling. 2017. “Semi-supervised classification with Graph Convolutional Networks.” International Conference on Learning R e presentations (ICLR), Toulon, France, arXiv preprint arXiv:1609.02907.

Kirillov, A., Y. Wu, K. He, and R. Girshick. 2020. “PointRend: Image Segmentation as Rendering.” IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 9796–9805. 10.1109/CVPR42600.2020.00982.

Kolesnikov, A., L. Beyer, X. Zhai, J. Puigcerver, J. Yung, S. Gelly, and N. Houlsby. 2020. “Big Transfer (Bit): General Visual Representation Learning.” European Conference on Computer Vision (ECCV) 491–507. https://arxiv.org/abs/1912.11370?context=cs.LG

Kouw, W. M., and M. Loog. 2018. “An Introduction to Domain Adaptation and Transfer Learning.” arXiv preprint. https://arxiv.org/abs/1812.11806

Li, X., F. Ling, G. M. Foody, Y. Ge, Y. Zhang, and Y. Du. 2017. “Generating a Series of Fine Spatial and Temporal Resolution Land Cover Maps by Fusing Coarse Spatial Resolution Remotely Sensed Images and Fine Spatial Resolution Land Cover Maps.” Remote Sensing of Environment 196: 293–311. doi:10.1016/j.rse.2017.05.011.

Li, D., D. Lu, N. Li, M. Wu, and X. Shao. 2019. “Quantifying Annual land-cover Change and Vegetation Greenness Variation in a Coastal Ecosystem Using Dense time-series Landsat Data.” GIScience & Remote Sensing 56 (5): 769–793. doi:10.1080/15481603.2019.1565104.

Ling, F., D. Boyd, Y. Ge, G. M. Foody, X. Li, L. Wang, Y. Zhang, et al. 2019a. “Measuring River Wetted Width from Remotely Sensed Imagery at the Subpixel Scale with a Deep Convolutional Neural Network.” Water Resources Research 55 (7): 5631–5649. doi: 10.1029/2018WR024136.

Ling, F., and G. M. Foody. 2019b. “Super-resolution Land Cover Mapping by Deep Learning.” Remote Sensing Letters 10 (6): 598–606. doi: 10.1080/2150704X.2019.1587196.

Liu, X., Y. Huang, X. Xu, X. Li, X. Li, P. Ciais, P. Lin, et al. 2020. “High-spatiotemporal-resolution Mapping of Global Urban Change from 1985 to 2015.” Nature Sustainability 3 (7): 564–570. doi:10.1038/s41893-020-0521-x.

Liu, X., F. Pei, Y. Wen, X. Li, S. Wang, C. Wu, Y. Cai, et al. 2019. “Global Urban Expansion Offsets climate-driven Increases in Terrestrial Net Primary Productivity.” Nature Communications 10 (1): 5558. doi:10.1038/s41467-019-13462-1.

Loveland, T. R., B. C. Reed, J. F. Brown, D. O. Ohlen, Z. Zhu, L. Yang, J. W. Merchant, et al. 2000. “Development of Global Land Cover Characteristics Database and IGBP DISCover from 1km AVHRR Data.” International Journal of Remote Sensing 21 (6–7): 1303–1330. doi:10.1080/014311600210191.

Maggioni, E., Y. Tarabalka, G. Charpillet, and P. Alliez. 2017. “Can Semantic Labelling Methods Generalize to Any City? The Inria Aerial Image Labeling Benchmark.” IEEE International Symposium on Geoscience and Remote Sensing, Fort Worth, TX, USA, 3226–3229. doi:10.1109/IGARSS.2017.8127684.

Ma, X., Y. Hong, Y. Song, and Y. Chen. 2019. “A Super-Resolution Convolutional-Neural-Network-Based Approach for Subpixel Mapping of Hyperspectral Images.” IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 12 (12): 49304939. doi:10.1109/JSTARS.2019.2941089.

Marin, D., Z. He, P. Vajda, P. Chatterjee, S. Tsai, F. Yang, and Y. Boykov. 2019. “Efficient Segmentation: Learning Downsampling near Semantic Boundaries.” IEEE International Conference on Computer Vision (ICCV), Seoul, Korea, 2131–2141. doi:10.1109/ICCV.2019.00222.

Maxwell, A. E., T. A. Warner, and L. Andrés Guillén. 2021. “Accuracy Assessment in Convolutional Neural network-based Deep Learning Remote Sensing studies—Part 2: Recommendations and Best Practices.” Remote Sensing 13 (13): 2591. doi:10.3390/rs13132591.

Mertens, K. C., B. D. Baets, L. P. C. Verbeke, and R. D. Wulff. 2006. “A Sub-pixel Mapping Algorithm Based on sub-pixel/pixel Spatial Attraction Models.” International Journal of Remote Sensing 27 (15): 3293–3310. doi:10.1080/01431160500497127.

Muad, A. M., and G. M. Foody. 2012. “Super-resolution Mapping of Lakes from Imagery with a Coarse Spatial and Fine Temporal Resolution.” International Journal of Applied Earth Observation and Geoinformation 15: 79–91. doi:10.1016/j.jag.2011.06.002.

Nguyen, Q. M., P. M. Atkinson, and H. G. Lewis. 2011. “Super-resolution Mapping Using Hopfield Neural Network with Panchromatic Imagery.” International Journal of Remote Sensing 32 (21): 6149–6176. doi:10.1080/01431161.2010.507797.
Pan, S., Y. Tao, C. Nie, and Y. Chong. 2020. “PEGNet: Progressive Edge Guidance Network for Semantic Segmentation of Remote Sensing Images.” IEEE Geoscience and Remote Sensing Letters 18 (4): 637–641. doi:10.1109/LGRS.2020.2983464.

Pekel, J. F., A. Cottam, N. Gorelick, and A. S. Belward. 2016. “High-resolution Mapping of Global Surface Water and Its long-term Changes.” Nature 540 (7633): 418. doi:10.1038/nature20584.

Poudyal, A. 2013. Spatial Statistics and Super Resolution Mapping for Precision Agriculture Using VHR Satellite Imagery [M]. Netherlands: University of Twente.

Shang, C., X. Li, G. M. Foody, Y. Du, and F. Ling. 2020. “Superresolution Land Cover Mapping Using a Generative Adversarial Network.” IEEE Geoscience and Remote Sensing Letters 19: 1–5. doi:10.1109/LGRS.2020.3020395.

Shannon, C. E. 1948. “A Mathematical Theory of Communication.” Bell System Technical Journal 27 (3): 379–423. doi:10.1002/j.1538-7305.1948.tb01338x.

Shi, W., J. Caballero, F. Huszar, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang. 2016. “Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network.” IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 1874–1883. doi:10.1109/CVPR.2016.207.

Song, M., Y. Zhong, A. Ma, and R. Feng. 2019. “Multiojective Sparse Subpixel Mapping for Remote Sensing Imagery.” IEEE Transactions on Geoscience and Remote Sensing 57 (7): 4490–4508. doi:10.1109/TGRS.2019.2891354.

Stehman, S. V., and G. M. Foody. 2019. “Key Issues in Rigorous Accuracy Assessment of Land Cover Products.” Remote Sensing of Environment 231: 111199. doi:10.1016/j.rse.2019.05.018.

Su, Y. 2019. “Integrating a scale-invariant feature of Fractal Geometry into the Hopfield Neural Network for super-resolution Mapping.” International Journal of Remote Sensing 40 (23): 8933–8954. doi:10.1080/01431161.2019.1624865.

Tateishi, R., B. Uriyangqai, H. Al-Bilbisi, M. A. Ghar, J. Tsend-Ayuush, T. Kobayashi, A. Kasimu, et al. 2011. “Production of Global Land Cover Data - GLCNMO.” International Journal of Digital Earth 4 (1): 22–49. doi:10.1080/17538941003777521.

Tong, X., G. Xia, Q. Lu, H. Shen, S. Li, S. You, L. Zhang, et al. 2020. “Land-cover classification with high-resolution Remote Sensing Images Using Transferable Deep Models.” Remote Sensing of Environment 237:111322. doi:10.1016/j.rse.2019.111322.

Tuia, D., C. Persello, and L. Bruzzzone. 2016. “Domain Adaptation for the Classification of Remote Sensing Data: An Overview of Recent Advances.” IEEE Geoscience and Remote Sensing Magazine 4 (2): 41–57. doi:10.1109/MGRS.2016.2548904.

Vu, T., H. Jain, M. Bucher, M. Cord, and P. Perez. 2019. “ADVENT: Adversarial Entropy Minimization for Domain Adaptation in Semantic Segmentation.” IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 2517–2526. doi:10.1109/CVPR.2019.00262.

Wang, Q., X. Ding, X. Tong, and P. M. Atkinson. 2021b. “Spatio-temporal Spectral Unmixing of time-series Images.” Remote Sensing of Environment 259: 112407. doi:10.1016/j.rse.2021.112407.

Wang, Q., W. Shi, P. M. Atkinson, and Y. Zhao. 2015. “Downscaling MODIS Images with area-to-point Regression Kriging.” Remote Sensing of Environment 166: 191–204. doi:10.1016/j.rse.2015.06.003.

Wang, Q., C. Zhang, X. Tong, and P. M. Atkinson. 2020. “General Solution to Reduce the Point Spread Function Effect in Subpixel Mapping.” Remote Sensing of Environment 251: 112054. doi:10.1016/j.rse.2020.112054.

Wang, P., G. Zhang, L. Wang, H. Leung, and H. Bi. 2019. “Subpixel Land Cover Mapping Based on Dual Processing Paths for Hyperspectral Image.” IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 12 (6): 1835–1848. 10.1109/JSTARS.2019.2910539. 2910539.

Wegener, M. 1990. “Destriping Multiple Sensor Imagery by Improved Histogram Matching.” International Journal of Remote Sensing 11 (5): 859–875. doi:10.1080/01431169008955060.

Wernick, I. K., P. Ciais, J. Fridman, P. Högborg, K. T. Korhonen, A. Nordin, P. E. Kauppi, et al. 2021. “Quantifying Forest Change in the European Union.” Nature 592 (7856): E13–E14. doi:10.1038/s41586-021-03293-w.

Xiao, J., and A. Moody. 2005. “A Comparison of Methods for Estimating Fractional Green Vegetation Cover Within a desert-to-upland Transition Zone in Central New Mexico, USA.” Remote Sensing of Environment 98 (2–3): 237–250. doi:10.1016/j.rse.2005.07.011.

Xiong, J., P. S. Thenkabail, J. C. Tilton, M. K. Gurna, P. Teluguamla, A. Oliphant, R. Congalton, et al. 2017. “Nominal 30m Cropland Extent Map of Continental Africa by Integrating Pixel-Based and Object-Based Algorithms Using Sentinel-2 and Landsat-8 Data on Google Earth Engine.” Remote Sensing 9 (10): 1065. doi:10.3390/rs9101065.

Xu, X., X. Tong, A. Plaza, J. Li, Y. Zhong, H. Xie, L. Zhang, et al. 2018. “A New Spectral-Spatial Sub-Pixel Mapping Model for Remotely Sensed Hyperspectral Imagery.” IEEE Transactions on Geoscience and Remote Sensing 56 (11): 6763–6778. doi:10.1109/TGRS.2018.2842748.

Yang, W., J. Feng, J. Yang, F. Zhao, J. Liu, Z. Guo, and S. Yan. 2017. “Deep Edge Guided Recurrent Residual Learning for Image Super-Resolution.” IEEE Transactions on Image Processing 26 (12): 5895–5907. doi:10.1109/TIP.2017.2750403.

Yin, Z., F. Ling, X. Li, X. Cai, H. Chi, X. Li, L. Wang, Y. Zhang, and Y. Du. 2021. “A Cascaded Spectral-Spatial CNN Model for Super-Resolution River Mapping with MODIS Imagery.” IEEE Transactions on Geoscience and Remote Sensing 60:5614213. doi:10.1109/TGRS.2021.3129789.

Yu, X., X. Wu, C. Luo, and P. Ren. 2017. “Deep Learning in Remote Sensing Scene Classification: A Data Augmentation Enhanced Convolutional Neural Network Framework.” GeScience & Remote Sensing 54 (5): 741–758. doi:10.1080/15481603.2017.1323377.
Zhang, C., Y. Chen, and D. Lu. 2015. “Detecting Fractional Land-Cover Change in Arid and Semiarid Urban Landscapes with Ulti-temporal Landsat Thematic Mapper Imagery.” GIScience & Remote Sensing 52 (6): 700–722. doi:10.1080/15481603.2015.1071965.

Zhang, X., Y. Ge, F. Ling, J. Chen, Y. Chen, and Y. Jia. 2021b. “Graph Convolutional networks-based super-resolution Land Cover Mapping.” IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 14: 7667–7681. doi:10.1109/JSTARS.2021.3100400.

Zhang, Y., F. Ling, G. M. Foody, Y. Ge, D. S. Boyd, X. Li, Y. Du, and P. M. Atkinson. 2019. “Mapping Annual Forest Cover by Fusing PALSAR/PALSAR-2 and MODIS NDVI during 2007-2016.” Remote Sensing of Environment 224: 74–91. doi:10.1016/j.rse.2019.01.038.

Zhang, X., L. Liu, X. Chen, Y. Gao, S. Xie, and J. Mi. 2021a. “GLC_FCS30: Global land-cover Product with Fine Classification System at 30 m Using time-series Landsat Imagery.” Earth System Science Data 13 (6): 2753–2776. doi:10.5194/essd-13-2753-2021.

Zhong, Y., Y. Wu, X. Xu, and L. Zhang. 2015. “An Adaptive Subpixel Mapping Method Based on MAP Model and Class Determination Strategy for Hyperspectral Remote Sensing Imagery.” IEEE Transactions on Geoscience and Remote Sensing 53 (3): 1411–1426. doi:10.1109/TGRS.2014.2340734.

Zhong, Y., Y. Wu, L. Zhang, and X. Xu. 2014. “Adaptive MAP sub-pixel Mapping Model Based on Regularization Curve for Multiple Shifted Hyperspectral Imagery.” ISPRS Journal of Photogrammetry and Remote Sensing 96: 134–148. doi:10.1016/j.isprsjprs.2014.06.019.

Zhu, Q., Y. Lei, X. Sun, Q. Guan, Y. Zhong, L. Zhang, and D. Li. 2022. “Knowledge-guided Land Pattern Depiction for Urban Land Use Mapping: A Case Study of Chinese Cities.” Remote Sensing of Environment 272: 112916. doi:10.1016/j.rse.2022.112916.

Zhu, Z., and C. E. Woodcock. 2014. “Continuous Change Detection and Classification of Land Cover Using All Available Landsat Data.” Remote Sensing of Environment 144: 152–171. doi:10.1016/j.rse.2014.01.011.